THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXIII January, 1942 Number 1

THE COMPARATIVE PERFORMANCE OF A GROUP OF OLD-DULL AND YOUNG-BRIGHT CHILDREN ON SOME ITEMS OF THE REVISED STANFORD-BINET SCALE OF INTELLIGENCE, FORM L

SAMUEL R. LAYCOCK AND STANLEY CLARK

College of Education, University of Saskatchewan

I. ORIGIN OF THE PROBLEM

The long-standing controversy as to the relative effects of "nature" and "nurture" seems currently to centre on the question of the degree to which favorable or unfavorable environments affect the developing intelligence of the child. The results of recent research and opinion have been summarized in the *Thirty-ninth Yearbook* of the National Society for the Study of Education.[1] Unfortunately no conclusion was reached on the point mentioned above.

The difficulty of settling the problem would seem to arise out of the fact that no test measures intelligence directly. Rather, a test purports to find an estimate of an individual's mental ability by measuring the outward manifestations of functioning intelligence. The value of an intelligence test seems to depend upon the relative freedom of its component items from the influence of special training and knowledge which are not available to the great majority of children. If a test is weighted with such items the resulting test score may give a somewhat spurious estimate of the testee's ability. Those psychologists who support Spearman's Two-Factor Theory of Intelligence, put forward the claim that test items which depend for success upon the mere reproduction of factual knowledge or upon the exercising of acquired skills are invalid for testing purposes. On the other hand, items involving eduction are the most desirable items. Test items

[1] *Thirty-ninth Yearbook.* The National Society for the Study of Education. Public School Publishing Company, Bloomington, Ill., 1940.

on the various revisions of the Binet Scale have been criticized[1] for their alleged susceptibility to environmental influences. Indeed, current controversy often resolves itself into a question not of whether environment affects the intelligence of the child, but whether environment affects the IQ as obtained by the Binet or some other scale of intelligence.

In order to get valid results with any intelligence scale (even a performance scale) the child tested must have had a background reasonably comparable to the average background of those children on whose performances the norms of the test were secured. It has long been known that, with tests like the Stanford-Binet Scale, a child who comes from a home where English is not spoken, is not adequately measured as to intelligence by these scales. The authors believe that children who have not had average opportunity at school cannot be adequately measured by this scale. Evidence for this is seen in the work of Gordon[2] with canal-boat and gypsy children. It would seem that the decline in IQ of these children is due not to a decline in the rate of intellectual growth, but to the fact that the measuring scale necessitates an average opportunity for academic education.

As one of the most widely used tests is the Revised Stanford-Binet Scale, it is imperative that studies be made of this test with a view to determining which items are largely reproductive (that is, the result of experience and training) and which are eductive. To the degree that reproductive items enter into the test it will necessarily show the effects of varying environments.

There are several methods of approaching this problem. One method of approach would be to give intensive training on material similar or identical to that of which the test items are composed. This has already been done with regard to the former Stanford-Binet scale.[3] A second method of approach could be made through an investigation into the cumulative effects of environment on a group of old-dull

---

[1] See Cattell, Psyche: "Nature, Nurture and Mental Growth." *The Review of Educational Research*, Vol. III, No. 2, April, 1933.

[2] Gordon, H.: *Mental and Scholastic Tests among Retarded Children.* London, Board of Education Pamphlet No. 44, 1923.

[3] Casey, Mary L., Davidson, Helen P., Harter, Doris I.: "Three Studies of the Effect of Training in Similar and Identical Material upon Stanford-Binet Test Scores." *Twenty-seventh Yearbook.* National Society for the Study of Education. Public School Publishing Co., Bloomington, Ill., 1928.

children as compared to a group of young-bright children of similar mental age. It would be well to know if, in the case of the older children, the fact that they had been exposed to environmental influences for a longer period of time affects their performance on specific items of the Revised Stanford-Binet as compared with the performance of the younger but brighter individuals who have been subjected to approximately the same general environmental conditions for a shorter time. It is the aim of the present study to investigate this problem by comparing the test responses on the Revised Stanford-Binet of a group of old-dull children matched with regard to mental age with a group of young-bright children.

II. THE SCOPE AND METHODS OF THE STUDY

The children chosen for this study were selected from the twelve public schools of the City of Saskatoon, Saskatchewan, Canada. The mental age range of the two groups was approximately eight years, six months, to ten years, six months.

The following criteria were used in selecting pupils for the old-dull group: (1) Chronological age, eleven to fifteen years; (2) IQ 85 and under; (3) They were chosen from the special classes for slow-learning children or (4) From the lists of the children recommended to the school psychologist for examination with a view to placement in the above classes.

Children for the young-bright group were chosen by the use of the following criteria: (1) Chronological age six to nine years; (2) IQ 115 and over; (3) They were picked for testing on the basis of teachers' judgments of their school performance and on the basis of test scores on a group mental test.

The children's home environment was measured by the Sim's Score Card for Socio-Economic Status, Form C. They were matched on home rating in so far as the scale measures it. No child was selected where English was not the language spoken at home.

In order to keep the personal equation constant all the children were given the Revised Stanford-Binet Scale, Form L, by the one examiner.

The specific test items considered were those on years VII to XII inclusive.

The old-dull and the young-bright were matched in pairs. The members of each pair were equated on mental age (not more than two months apart), sex, socio-economic status (as measured by the Sim's

Score Card and allowing a variation of one decile point) and on the general locality in which they lived.

The final groups consisted of thirty-four boys and six girls each. Sixteen pairs were matched in their own schools and over half in the same or adjacent areas.

The distribution of mental ages in the two groups is found in Table I.

TABLE I.—THE MEAN, THE STANDARD-DEVIATION AND THE DISTRIBUTION OF MENTAL AGES IN THE OLD-DULL AND THE YOUNG-BRIGHT GROUPS

MA months	Old-dull	Young-bright
127–131	2	1
123–127	2	3
119–123	8	9
115–119	6	5
111–115	3	3
107–111	6	5
103–107	10	11
99–103	3	3
Total	40	40
Mean	113.1	113.0
SD	8.02	8.00

The table giving the distribution of the decile ratings for socio-economic status for each group will not be included here. Statistical treatment showed that there were no significant differences between the means of the two groups. According to Sim's groupings of the classification of parental occupations, the majority of the fathers of the subjects were located in the lower brackets.

The Variable Factor.—The purpose of this study is to investigate the effects on Stanford-Binet performance of having lived longer in a comparatively similar environment. The variable factor is chronological age. Other factors such as length of time at school and the school grade of the subjects are more or less contingent upon this. The actual age-range of the old-dull subjects was from 11-4 to 14-9 and the children were evenly distributed over this range. The mean chronological age of this group was 154.70 months with a Standard Deviation of 11.64. The age range of the young-bright children was from 6-2 to 8-10. These children were not evenly distributed through this age

range. Over half had an age between eighty and ninety-six months. The Mean Age was 92.60 months with a Standard Deviation of 6.68.

The members of the old-dull group have, on an average, lived 62.1 months or five years, 2.1 months longer than the younger but brighter children. The differences between matched children varied from seven years, five months to three years, nine months.

With respect to grading, ten of the old-dull children were in grade III, eight in grade IV, six in grade V, two in grade VI and the remaining fourteen were in the special classes. Of the brighter group, eight were in grade I, twenty-six in grade II, four in grade III and two in grade IV.

The duller children had been in school for an average period of five years, eight months; the younger children, for an average of one year, five months.

An examination of the IQ distributions of the groups indicates a mean difference of almost 50 IQ points. The Mean IQ of the old-dull group was 73.75 with a standard deviation of 5.32. The range in IQ for this group was from 64 to 85. The young-bright children ranged in IQ from 115 to 144 with a Mean of 122.60 and a Standard Deviation of 6.29.

III. THE RESULTS

Table II gives the record of successes on each test item on the Revised Stanford-Binet for both groups of pupils. There was no differentiation between the groups at the seven-year level as the items were sufficiently easy for practically all children to pass.

There is a general tendency for the younger children to be slightly superior on tests of immediate memory, *e.g.*, in Memory for Sentences, VIII:6 and XI:4; in Memory for Digits in X:6; in Repeating Digits Reversed in IX:6 and XII:4; in the Memory for Words in XIII:2; and in the Memory for Designs, IX:3. The duller children were somewhat better than the brighter on the vocabulary tests and with the abstract words. The two groups are fairly even on the first two verbal absurdity tests, VIII:3 and IX:2, but the older children excel on the higher levels, XI:2 and XII:2. The largest percentage difference of successes in favor of the older group occurred in item X:2, Picture absurdities II, where a white man is attacked by two Indians while he chooses to shoot at a third Indian some distance away.

The old-dull did better in making change, IX:5, there being a twenty per cent superiority. Only two of the children failed to get two of the three items. They were also more successful in figuring

out Donald's dilemma on XI:5. It was noticeable that, whereas the old-dull were right or wrong immediately on this item, some of the brighter children were inclined, in giving possible reasons for the burning of the lad's clothing, to come upon the answer by a process of reasoning. The latter group appeared to rely on eduction and the former on reproduction.

Twelve of the older group and thirteen of the younger passed the Plan of Search, XIII:1. This item appears to be relatively too easy for these children. Krugman[1] found this to be the case among New York subjects.

Paper-cutting I did not discriminate between the two groups on the nine-year level, but the brighter children were twice as successful on this test in year XIII.

The brighter children were slightly superior on the tests involving similarities, VIII:4 and XI:6.

In spite of spending an average of some four years longer in school, the old-dull had four more failures than the young-bright on Reading and Report, X:3.

There was a difference of only two successes in favor of the old-dull on the free association test, X:5. Those who passed in both groups named an average of 32.5 words.

The young-brights had a somewhat superior performance on Memory for Stories VIII:2, while the older group did better on VIII:5, Comprehension IV. There was little or no difference in results on the items Verbal Absurdities I, VIII:3, Paper-cutting I, IX:1, Verbal Absurdities II, IX:2, Finding Reasons I, X:4, Word Naming X:5, Memory for Designs, XI:1, and Plan of Search, XIII:1. Success on the items above XIII:3 was too rare to warrant discussion.

IV. THE INTERPRETATION OF RESULTS

In order to obtain some idea of the significance of these percentage differences, it was decided to use the formula for the standard error of the difference between two percentages.[2] This formula is $\sigma D\rho = \sqrt{\frac{p_1q_1}{N_1} + \frac{p_2q_2}{N_2}}$ where $\sqrt{\frac{p_1q_1}{N_1}}$ and $\sqrt{\frac{p_2q_2}{N_2}}$ are the standard errors of the old-dull and the young-bright percentages; p stands for the proportion passing and q for the proportion failing the particular item. By

[1] Krugman, Morris: "Some Impressions of the Revised Stanford-Binet Scale." *Journal of Educ. Psychology*, Vol. XXX, No. 8, Nov. 1939, pp. 594–603.

[2] Garrett, H. E.: *Statistics in Psychology and Education.* Longmans, Green and Co., New York, 1939.

TABLE II.—THE DESCRIPTION OF EACH ITEM, THE NUMBER AND PERCENTAGE OF OLD-DULL AND YOUNG-BRIGHT CHILDREN PASSING EACH ITEM, AND THE PERCENTAGE OF DIFFERENCE IN FAVOR OF THE OLD-DULL OR YOUNG-BRIGHT GROUP

Year	Description of item	Old-dull successes		Young-bright successes		Percentage difference in favor of	
		Number	Per cent	Number	Per cent	*O-D*	*Y-B*
VII							
1	Picture absurdities	40	100.0	40	100.0		
2	Similarities: Two things	39	97.5	37	92.5	5.5	
3	Copying diamond	39	97.5	40	100.0	....	2.5
4	Comprehension III	40	100.0	39	97.5	2.5	
5	Opposite analogies	39	97.5	40	100.0	....	2.5
6	Repeating five digits	37	92.5	38	95.0	....	2.5
VIII							
1	Vocabulary	38	95.0	34	85.0	10.0	
2	Memory for stories	32	80.0	37	92.5	....	12.5
3	Verbal absurdities I	29	72.5	29	72.5		
4	Similarities and differences	30	75.0	35	87.5	....	12.5
5	Comprehension IV	33	82.5	27	67.5	15.0	
6	Memory for sentences III	28	70.0	31	77.5	....	7.5
IX							
1	Paper-cutting I	29	72.5	31	77.5	....	5.0
2	Verbal absurdities II	24	60.0	23	57.5	2.5	
3	Memory for designs	22	55.0	28	70.0	....	15.0
4	Rhymes: new form	25	62.5	28	70.0	....	7.5
5	Making change	38	95.0	30	75.0	20.0	
6	Repeating four digits reversed	22	55.0	26	65.0	....	10.0
X							
1	Vocabulary	17	42.5	11	27.5	15.0	
2	Picture absurdities II	32	80.0	23	57.5	22.5	
3	Reading and report	11	27.5	15	37.5	....	10.0
4	Finding reasons I	18	45.0	17	42.5	2.5	
5	Word naming	18	45.0	16	40.0	5.	
6	Repeating six digits	15	37.5	21	52.5	....	15.0
XI							
1	Memory for designs	14	35.0	14	36.0		
2	Verbal absurdities III	12	30.0	5	12.5	17.5	
3	Abstract words I	10	25.0	6	15.0	10	
4	Memory for sentences IV	11	27.5	17	42.5	....	15.0
5	Problem situation	14	35.0	9	22.5	12.5	
6	Similarities: Three things	10	25.0	13	32.5	....	7.5
XII							
1	Vocabulary	4	10.0	1	2.5	7.5	
2	Verbal absurdities II	12	30.0	5	12.5	17.5	
3	Response to pictures II	0	0	0	0	....	
4	Repeating five digits reversed	3	7.5	7	17.5	....	10.0
5	Abstract words II	4	10.0	1	2.5	7.5	
6	Minkus completion	1	2.5	0	0	2.5	
XIII							
1	Plan of search	12	30.0	13	32.5	....	2.5
2	Memory for words	1	2.5	6	15.0	....	12.5
3	Paper-cutting I	7	17.5	14	35.0	....	17.5
4	Problem of fact	3	7.5	3	7.5		
5	Disconnected sentences	0	0	2	5.0	....	5.0
6	Copying bead chain from memory	3	7.5	1	2.5	5.0	
XIV							
2	Induction	0	0	1	2.5	....	2.5
3	Picture absurdity III	2	5.0	2	5.0		
5	Orientation: direction I	0	0	1	2.0	....	2.5
SA II							
3	Repeating eight digits	0	0	2	5.0	....	5.0

dividing the observed difference by the standard error of the difference the critical ratio is obtained. It is customary to consider a critical ratio of 3 as virtual certainty that the true difference is greater than

zero. Guilford[1] states that a critical ratio between 2.00 and 3.00 indicates that there is fair certainty that the true difference is greater than zero. Table III gives the standard errors for the differences. The disadvantage in using the standard error of a difference formula in this study lies in the fact that there is only a small number of cases (forty) and the difference will have to be large to show up significantly.

It will be seen from Table III that there are no critical ratios of three or more. This would indicate that there is no high degree of certainty that, if a larger number of cases had been used in this investigation, the difference would still favor the same group. Obviously the environmental influences to which the old-dull children have been subjected on an average of five years longer than the young-bright have not caused any outstandingly significant differences so far as these are shown by the statistical formulas. Further study with other groups and with larger numbers will be needed before definite conclusions are reached.

With two test items there is fair certainty that the old-dull children will surpass the young-bright and that environmental factors have affected test response. These are IX:5, Making change and X:2, Picture absurdities II. The item consisting of the arithmetical operation of making change has been criticized in this regard. After attending school for an average period of five years, six months the older children must have had a great deal of drill on this type of question. In addition, being dull will probably have resulted in their getting more attention than the normal child. A further indication of the influence of the school in solving these questions is given in an analysis of the failures among the young-bright. The average chronological age of those children from this group who passed the test was 94.7 months; of the failures, 85.8 months. Seven of the eight pupils from grade I and three from grade II who failed had an average IQ of 125.1 compared to 121.8 for the successful brights. The failures were somewhat brighter than the rest of the group but apparently had not, due to shorter time in school, adequate opportunity to acquire the skills necessary to answer the question. In Picture absurdities II there is some evidence that this response is conditioned by life experience.

The older children tend toward superiority over the younger children of the same mental age on the vocabulary item. It is possible

[1] GUILFORD, J. P.: *Psychometric Methods.* McGraw-Hill Book Co., Inc., New York, 1936.

TABLE III.—THE PERCENTAGE DIFFERENCE, THE STANDARD ERROR OF THE DIFFERENCE AND THE CRITICAL RATIO FOR EACH ITEM ON THE STANFORD-BINET, FORM L, YEARS VIII TO XIII

Item	Percentage difference in favor of old-dull	Percentage difference in favor of young-bright	Standard error of the difference	Critical ratio
VIII: 1	10.0	....	6.6	1.51
2	....	12.5	7.5	1.67
3	0	0	0	
4	....	12.5	8.6	1.45
5	15.0	....	9.5	1.58
6	....	7.5	9.8	.77
IX: 1	....	5.0	9.6	.52
2	2.5	....	11.0	.23
3	....	15.0	10.6	1.41
4	....	7.5	10.5	.71
5	20.0	....	7.7	2.52
6	....	10.0	11.3	.88
X: 1	15.0	....	10.5	1.43
2	22.5	....	10.0	2.25
3	....	10.0	10.4	.96
4	2.5	....	11.1	.23
5	5.0	....	11.0	.45
6	....	15.0	11.0	1.36
XI: 1	0	0	0	
2	17.5	....	8.9	1.96
3	10.0	....	8.9	1.13
4	....	15.0	10.5	1.43
5	12.5	....	10.0	1.25
6	....	7.5	10.0	.75
XII: 1	7.5	....	5.3	1.41
2	17.5	....	8.9	1.96
3	0	0	0	
4	....	10.0	5.3	1.88
5	7.5	....	5.3	1.41
6	2.5	....	2.4	1.04
XIII: 1	....	2.5	10.3	.24
2	....	12.5	6.2	2.01
3	....	17.5	9.6	1.82
4	0	0	0	
5	....	5.0	3.5	1.43
6	5.0	....	4.8	1.04

that, through having lived some five years longer, the latter will have acquired a larger vocabulary in spite of their dullness. An analysis of the total vocabulary score, disregarding year levels, indicates that the old-dull averages 10.3 words and the young-bright group 9.37 words. The standard deviations are 2.06 and 1.85, respectively. The standard error for the difference between the means is .44 and the critical ratio 2.11. There is, therefore, fair certainty that the true difference between the groups in vocabulary is greater than zero.

The Verbal absurdities on the eight- and nine-year levels show little differentiation between the groups. On the eleventh- and twelfth-year levels, however, there is a noticeable difference in favor of the dull children. Failure on item XI:2 on the part of the younger children was largely due to inability to understand the language used in sub-items (*b*) and (*c*). Fifteen were successful on part (*a*) but only two and three on parts (*b*) and (*c*). The old-dull successes were twenty, five and eight, respectively. Credit was given for XII:2 when the child made four or five concrete responses to the five sub-items on IX:2. The old-dull were more successful on part (*d*) presumably from having studied the story of Christopher Columbus. Success by the younger ones was more incidental; one pupil had read about Columbus in a book belonging to an older member of the family, while another boy who had spent a short time in an ungraded country school had heard the teacher taking up the story with one of the upper grades. Columbus is probably better known among the children in the United States where the test was standardized. The younger children were more successful on part (*e*)—not because they had studied the Gulf Stream, but by observing the inconsistency in the idea of seeing icebergs after they had melted. It would seem that Tests of Absurdities may, according to their content, be testing either eduction or reproduction.

There seem to be some grounds for believing that life experience has had some effect on test responses in regard to Comprehension IV. Situations are presented here in which it is conceivable that, having spent an average of five years longer in the social milieu, the older children will be more likely to have acquired the information necessary to answer the questions. This test would seem to be of a reproductive rather than an eductive type.

There have been discussed six types of items in which the old-dull children have demonstrated a marked, if not statistically sig-

nificant, degree of superiority over the younger but brighter pupils. The differences in regard to the groups included in this study might be attributable to environmental effects. As the groups were matched on mental age, it remains to reveal how the young-bright were able to make up these differences in order to obtain similar test scores. The items in which they were most significantly superior were Memory for Stories, Similarities and Differences, Memory for Sentence IV, Memory for Designs and Repeating six Digits. The most striking thing about their performance is the consistency with which they excel in tests of immediate memory. There seems to be some reason here for disagreement with those individuals who advocate the elimination of tests of immediate memory from tests of general intelligence. Success in tests of immediate memory would seem to depend on eductive processes—the ability to see meaningful relations in the material presented. This is true of sentences and words. It is probable, too, that in the tests for repeating digits, the brighter pupils who depend on eduction are able to throw the digits into patterns or groups in order to retain them.

One very significant fact appeared in the Memory for Stories, year VIII. The first question to be answered after the story of *The Wet Fall* has been read is to tell the name of the story. Twenty-eight young subjects contrasted with eight older children were able to give the correct title. The critical ratio between the percentage differences was 5.2 (statistically significant). Is this ability on the part of the bright children to reproduce the title due to the "insight" which results from greater eductive power—the power to see meaningful relations in the selection as a whole?

The superiority of the young children was not as significant on the item Reading and report, x:3. However, the fact that they did better than the older children is somewhat startling. The success of the young-bright would probably have been much greater had the younger ones had the greater opportunities to master the mechanics of reading. All the eight children from grade I failed because of this. The twenty-five young failures averaged six months behind the fifteen young successes in length of time at school, nine months behind in grading, and nine months behind in chronological age but were equal in IQ. It would seem that success on this item comes rapidly once children have acquired the necessary skills. The success of the young-bright in this test was apparently due to their eductive ability to see meaningful relations in the material read.

There were no successes on Response to Pictures II (XII:3). This test is a striking example of the degree to which many test items depend upon the testees having a background similar to that of the children on whose performance the norms are based. Extensive clinical experience suggests that this test is quite unsuitable to Canadian children. While the words "Western Union" may be nearly synonymous with "telegram" for American children, they have no such significance for Canadian children.

V. SUMMARY AND CONCLUSIONS

(1) Forty pairs of old-dull and young-bright children matched as to mental age, sex, and socio-economic status, were compared as to their performance on the items of the Revised Stanford-Binet Scale of Intelligence, Form L.

(2) While no statistically significant differences between the two groups were found, there was fair certainty in some cases that the true difference is greater than zero. This was true in favor of the old-dull with regard to test IX:5 (making change) and X:2 (Picture absurdities II) and in favor of the young-bright group with regard to XIII:2.

(3) The old-dull did better to the extent of ten per cent or more on Picture Absurdities II, Making Change, Verbal Absurdities III, Verbal Absurdities II, Comprehension IV, Vocabulary, Problem Situation, Abstract Words I.

(4) The young-bright group did better to the extent of ten per cent or more on Paper-cutting I, Memory for Sentences IV, Repeating six Digits, Memory for Designs, Memory for Words, Similarities and Differences, Memory for Stories, Repeating five Digits Reversed, Reading and Report, Repeating four Digits Reversed.

(5) There are indications that some tests of the Revised Stanford Binet are tests in which the psychological process of Reproduction plays a prominent part whereas other tests would appear to be highly Eductive in character.

(6) The policy of accepting an obtained IQ as valid without considering the extent to which the testee's background of environment and training differs from that of the children on whom the norms were obtained, may be called into question.

(7) Further research is needed to determine whether an obtained IQ should be regarded as a valid measure of intelligence when the child's environment has been that of a poverty-stricken orphanage or of an enriched nursery-school environment.

A STUDY OF THE INTEREST-ATTITUDE TEST SCORES OF DELINQUENT AND NON-DELINQUENT BOYS

M. E. ODOROFF

Division of Social Welfare, St. Paul, Minn.

AND

DALE B. HARRIS

Institute of Child Welfare, University of Minnesota

"Affective maturity," a concept of wide use among psychologists, does not have as generally accepted meaning as has "intellectual maturity." Although the level of intellectual maturity may be inferred from mental age obtained by tests, no analogous "emotional age" has been developed, nor have tests been available for determining it, largely because acceptable criteria of emotional development have been so difficult to secure.

As with most descriptive terms, emotional maturity has a variety of meanings. In some contexts it refers to the restraint of emotion, a concept that implies simply that the overt expression of emotion is undesirable, and hence to be controlled. In other contexts it refers to a quality of emotional responsiveness which emerges with increasing age; that is, with growth from infancy to adulthood, emotional responses become increasingly more typical of adjusted adulthood.[7] Occasionally the term refers to emotional behavior that conforms to the demands of a given social or cultural milieu; in order to be considered mature, the individual's behavior must be acceptable within the given society. From these examples of the meanings of affective maturity, it is clear that no single definition has as yet received widespread acceptance.

Prescott[7] has suggested that no one formulation of the concept is possible. His assertion is based on the premise that emotionally mature patterns of behavior are dependent on the organisms' needs in different situations. Such a premise makes it impossible to trace from any particular pattern of stimulation any given pattern of response. This premise, if accepted, makes untenable the test approach to the problem, since with different situations and changing needs of the organism the response at any one time is unique. On the other hand, our educational system proceeds rather effectively on the assumption of general similarity in situations which people meet, and in the intellectual needs which people develop. It may not exceed reason to

assume that environment makes similar emotional demands on many individuals, and that similar emotional needs may exist for smaller or larger cultural groupings. Although Prescott does not list them in his bibliography, he seems to have considered a variety of attempts to measure quantitatively an emotional maturity which increases with age in normal individuals. At any rate, he concludes[7] that "scales" or "sequence schedules" or "tests" purporting to measure affective maturity as a whole are predestined to failure by the very nature of the construct they are trying to measure.

Yet the efforts to seek a quantitative expression of emotional maturity have not been wholly fruitless. Furfey's scale for measuring developmental age,[3] concerned as it is with non-intellectual aspects of general behavior, undoubtedly incorporates the idea of emotional growth. Chambers,[1] working with the old Pressey *X*-0 test, developed a measure of emotional maturity. Weber[13,14] has produced a test device for obtaining an "emotional age" score. Willoughby's rating scale[15] is designed to express the degree to which an adult is free from immature emotional attitudes.

In 1933 the Presseys[8] published a revision of their earlier test, calling the new form the "Interest-Attitude Tests." As age increases, an increasing "maturity of emotional behavior" presumably shows itself in what an individual considers wrong, what he worries about, what he is interested in, and the traits he admires in people.

The construction of the test involved "selection (from earlier data) of those words which appeared for both sexes indicative of maturity, that is, words that were marked distinctly more often by either younger or older children."[8] Two categories of words are included in each subtest—words marked most frequently by younger individuals (scored plus), and words marked most frequently by older individuals (scored minus). The total score consists of an algebraic summation of plus and minus totals and decreases with age. "Emotional age" norms for each age and half age from eight to twenty-three are given by the Presseys based on scores of large samples of unselected boys and girls. The Presseys note that until there has been an opportunity "to follow up and investigate cases in some detail, the soundness and significance of such a concept of 'emotional age' will be uncertain. However, the concept would appear to have some merit."*

Durea,[2] by a comparison of the Interest-Attitude scores of delinquents with norms for corresponding age groups of unselected boys, has

* Manual for the Pressey Interest-Attitude Tests.

afforded a basis for the conclusion that delinquents are consistently retarded, age for age. This retardation is referred to as "emotional retardation."

THE PROBLEM

The present paper seeks to verify Durea's conclusions by comparing delinquent boys with a non-delinquent group more similar to delinquents in socio-economic and intellectual status than unselected boys, the basis of Durea's comparison. It also includes a study of the effect on Pressey Interest-Attitude scores of increasing mental maturity.

PROCEDURE AND RESULTS

The Pressey Interest-Attitude tests were administered in 1937 and 1939 to four hundred twelve boys of the Minnesota State Training School at the time of entrance. The subjects ranged in chronological age from 14-0 to 18-7. A few Negro and Indian boys were included—racial groups not represented in Durea's study. Mental age for all subjects were obtained from the new Kuhlmann Tests of Mental Development[5] administered by one of the writers, then psychologist for the Minnesota State Board of Control. The mental age value for each case was obtained within two months of the time the subject took the Pressey test. To eliminate all time discrepancies, however, Hilden's Table of Per cent of Average Development[4] was used to adjust the mental age values to an estimated value as of the actual date of the Pressey test. All chronological ages were likewise computed as of this date. Thus, a real effort was made to control time discrepancies, so often an uncontrolled factor in studies of large groups when tested at different times.

The mean intelligence of the Minnesota delinquents, represented by an IQ of 83 (using sixteen years as the maximum denominator)* or 88, using fifteen, as the maximum, is typical of institutionalized delinquents.[6] Delinquents, as a group, are retarded mentally. The question at once may be raised: "Is so-called 'emotional retardation' related to the commonly observed mental inferiority of delinquent boys?"

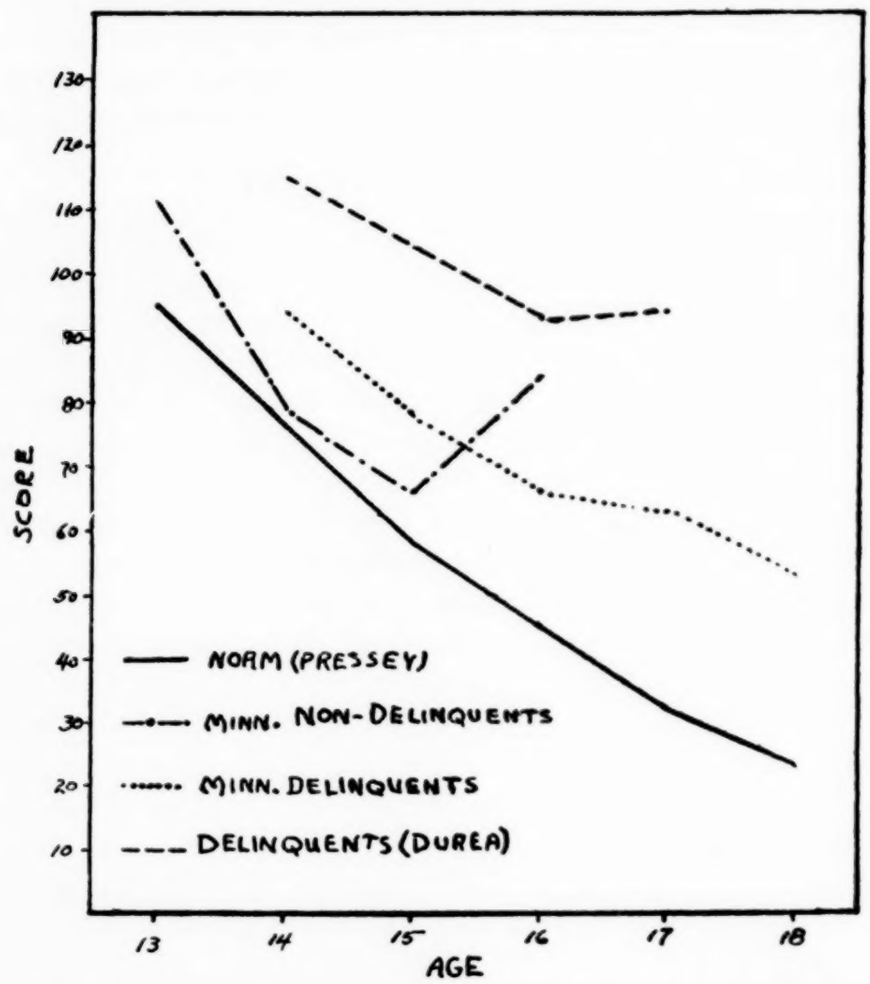
Table I and Fig. 1 compare the performance of the Minnesota delinquents with the group reported by Durea. While the Minnesota sample is emotionally retarded with respect to Pressey's normative values, this retardation is not so extensive as in the Ohio group.

* Kuhlmann Tests of Mental Development.

(Higher numerical scores are indicative of immaturity, according to Pressey's formulation.)

To check further on the emotional retardation of delinquent boys, samples of thirteen-, fourteen-, fifteen-, and sixteen-year-olds were obtained by testing the available boys of these ages in the seventh,

FIG. 1



eighth, and ninth grades of a St. Paul, Minnesota, junior high school in the Spring of 1940. This school is located in what may be termed a "delinquency area." While a few of the subjects may have had contact with the courts, the group as a whole was presumed to be non-delinquent. Since delinquent boys are not representative of boys in general in terms of intelligence, school achievement, or socio-economic status, this sample of non-delinquents from a delinquency area may be considered a better group for comparison than an unselected group. The school has the lowest average intelligence rate per class in St.

Paul (the mean Otis IQ for all boys studied was 95); the fathers of the boys were for the most part semi-skilled, unskilled, and unemployed workers. About thirty boys in the group could be placed in the skilled class of the Minnesota Occupational Scale. None came from classes I or II of this scale.

TABLE I.—INTEREST-ATTITUDE TESTS PERFORMANCE OF UNSELECTED AND SPECIAL SUBJECTS WHEN GROUPED ACCORDING TO CHRONOLOGICAL AGE

Chronological age group	Norm (unselected) median	Durea (delinquent)		Minnesota (delinquent)		Minnesota (non-delinquent)	
		N	Median	*N*	Median	*N*	Median
13	95.0	...	.....	...	.....	54	111.25
14	76.0	66	115.00	91	93.85	77	78.18
15	58.0	75	103.75	120	78.00	62	65.83
16	45.0	112	93.00	96	65.79	31	84.00
17	32.0	63	93.75	70	63.33		
18	23.0	...	.....	35	53.54		

TABLE II.—MEDIAN INTEREST-ATTITUDES TESTS SCORES OF DELINQUENT AND NON-DELINQUENT BOYS WHEN GROUPED ACCORDING TO MENTAL AGE*

Mental age group	Minnesota samples			
	Delinquents		Non-delinquents	
	N	Median	*N*	Median
9	41	87.5	20	170.0
10	53	87.3		
11	50	92.9	36	98.0
12	71	78.5	44	97.4
13	53	66.7	25	80.0
14	37	57.1	38	78.0
15	29	56.7	25	56.7
16	27	65.0	19	46.7
17	51	47.1		

* Delinquents—Kuhlmann Tests of Mental Development.
Non-delinquents—Mental age equivalents of scores on the Otis Advanced *S-A* exam.

When compared with unselected boys from a typical delinquency area, delinquents reveal less "emotional retardation" than when com-

pared to unselected boys in general. Table I shows that at age fourteen the sample of non-delinquents does not differ from Pressey's norms, while the non-delinquent sixteen-year-olds show even more retardation than delinquents of the same age. Usually the sixteen-year-olds available in the junior-high grades represent a selection of slightly duller than average boys of the same age. In part, this may explain the reversal of the trend in Pressey score in Table I and Fig. 1. In general, however, the definite trends of mean score with chronological age in both delinquent and non-delinquent groups support Pressey's findings.

TABLE III.—TEST PERFORMANCE OF DIFFERENT IQ GROUPS WITHIN EACH LIFE AGE LEVEL, MINNESOTA DELINQUENT BOYS

Chronological age group	Intelligence level							
	Below 75		75–89		90–99		100 and above	
	N	Median	*N*	Median	*N*	Median	*N*	Median
14	24	102.5	31	96.0	16	115.0	20	74.0
15	41	94.3	34	98.0	15	66.7	30	56.2
16	25	80.0	35	68.6	11	60.0	25	55.0
17–18	31	108.0	41	56.4	12	37.5	21	47.5

TABLE IV.—TEST PERFORMANCE OF DIFFERENT IQ GROUPS WITHIN EACH LIFE AGE LEVEL, MINNESOTA NON-DELINQUENT BOYS

Chronological age group	Intelligence level					
	Below 90		90–99		100 and above	
	N	Median	*N*	Median	*N*	Median
13	7	110.00	18	125.0	19	100.0
14	26	147.5	17	80.0	34	61.2
15	25	93.3	17	86.7	20	50.0
16	17	93.3	12	86.0	12	15.0

Is "emotional age" a function of general mental ability? The data of Table II indicate that such is the case to some degree. Tables III and IV indicate an appreciable difference in median scores for the extreme groups—the dull and the bright—when chronological age is held constant. Furthermore, within any one level of brightness, there

is not a consistent tendency toward increased median score with successive chronological age levels—it is actually a smaller and less stable tendency than might be supposed to exist from the general trend of the scores with age groups unselected for level of intelligence. This tendency gives additional confirmation that maturity as defined by the Pressey Tests is not wholly independent of mental maturity.

It should be noted also that the emotional age values for the median Pressey scores correspond more closely to the median scores for the chronological age levels than to the mental age levels in both the delinquent and non-delinquent groups. This finding for dull boys contrasts with Thorndike's findings for very bright children of eleven years median chronological age.[12] He found that the age equivalent on the Pressey Test falls much closer to mental age than to chronological age. In both instances the age equivalent for median Pressey scores falls between CA and MA. Thus, additional evidence appears that mental as well as chronological age effects performance on the Pressey Test.

DISCUSSION

Apropos of the empirical meaning of these tests, some additional observations can be made. It is, of course, true that reading ability develops concomitantly with mental maturity. While Pressey gives normative values beginning at emotional ages eight and one-half for boys and eight for girls, and grade norms start at the fifth, the test itself contains many polysyllabic words, which at first appearance seem quite difficult for youngsters of elementary-school age. The Thorndike *Teachers' Word Book*[10] provides one means of estimating the unusualness (in terms of frequency) of the words used. Table V gives an analysis of the terms used in the various tests, according to Thorndike's list. While this is only a crude check on the commonness or unusualness of the words used in the test, it is clear that those terms which the authors of the tests have found to indicate maturity are distinctly less common in general usage than those words characteristic of immature emotional attitudes. It is probable that young children or dull children with limited recognition vocabularies might have some difficulty in interpreting even some of the words significant for immaturity. Assuming that a child checks only words that he knows and that his comprehension is limited to the level of the first thousand words on Thorndike's list, he could obtain a score of 52 by checking each of the eighty-six words at this level. This score corresponds to

an emotional age of sixteen years. If he checked only immature words, his maximum score would be sixty-nine, or an emotional age of fifteen years, whereas if he checked only mature items his maximum score would be −17, or an emotional age of twenty years. Thus, a respectable degree of maturity could be obtained by a child of limited reading ability who faithfully carried out the test instructions, "to mark everything which you think is wrong" etc., and did not react to words he could not comprehend.

TABLE V.—POSITION OF THE STIMULUS WORDS OF THE INTEREST-ATTITUDE TEST IN THORNDIKE'S 10,000 WORD LIST

Position in Thorndike list	Words scored "plus" (immaturity)		Words scored "minus" (maturity)	
	N	Per cent	*N*	Per cent
1–1,000	69	27	17	17
1,000–5,000	125	48	38	38
5,000–10,000	47	18	36	36
Not in list	19	7	9	9
Total	260	100.0	100	100.0

Robert Thorndike[11] has indicated that the total score of the Pressey Test may depend on the subject's tendency to "check freely" both mature and immature items. His argument is that since there are two hundred sixty items out of a possible three hundred sixty in the test that are indicative of immaturity, the test gives an individual who "checks freely" more than twice as many chances to check immature as mature items. In reply, Pressey states that this tendency to check many rather than few items is probably an important expression of immaturity.[9] Table VI and Fig. 2 clearly show that in this study the one hundred mature items contribute little to the total maturity scores of the subjects. Increase in maturity of total score is brought about by a reduction in the number of immature words checked. If this finding should be confirmed generally, one could question the usefulness of the one hundred mature items in the test.

Apparently older and younger individuals react differently (by worrying, admiring, being interested, and the like) to the same verbal

symbols as well as to different classes or categories of symbols. This differential reaction would imply that attitudes and value concepts change as individuals mature intellectually and chronologically. At the present time one of the writers is engaged in a study to clarify this phase of the problem.

TABLE VI.—MEDIAN IMMATURITY, MATURITY, AND TOTAL SCORE BY CHRONOLOGICAL AGE FOR DELINQUENT AND NON-DELINQUENT GROUPS

Age	*N**	Median of plus items (immaturity)	Median of minus items (maturity)	Median total score
		Delinquents		
14	38	135.0	35.0	97.5
15	53	108.0	35.0	69.0
16	55	98.8	33.6	65.6
17–18	54	98.8	37.5	56.2
		Non-delinquents		
13	54	143.8	50.8	111.2
14	77	130.0	47.0	78.2
15	62	104.2	36.8	65.8
16	31	127.5	42.5	84.0

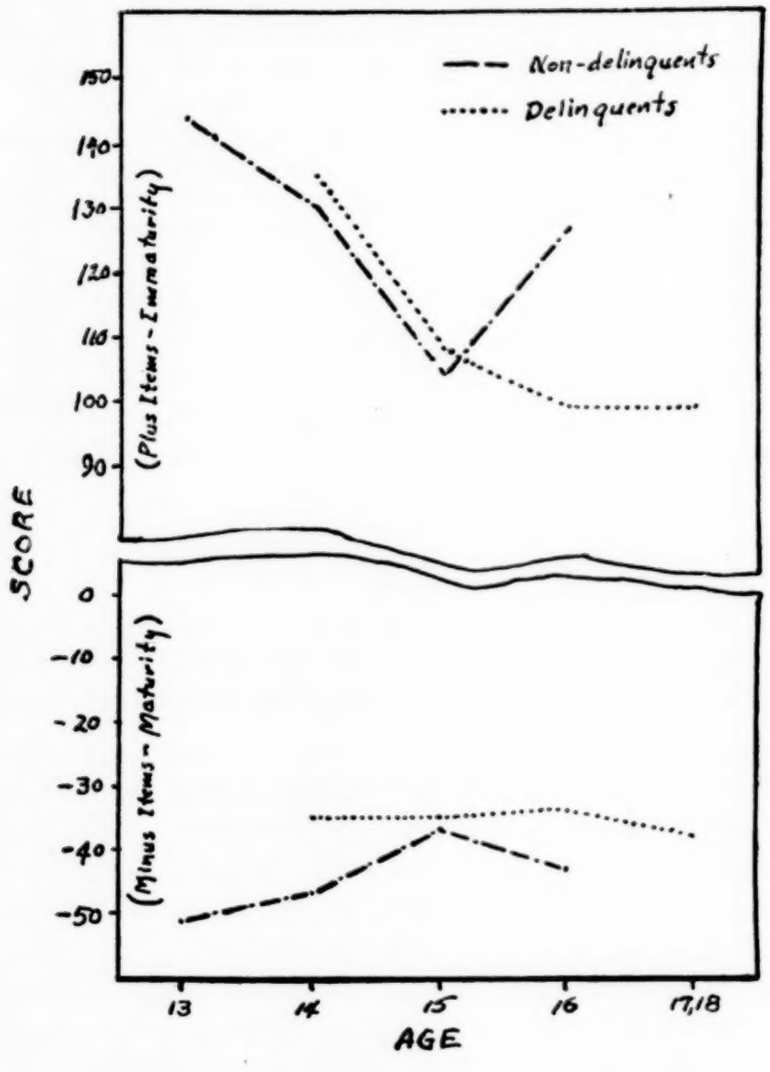
* These numbers are smaller than those indicated in Table I because subscores for two groups of tests were not available at the time these data were summarized.

It has been the writers' observations that some boys unable to read specific words have checked them anyway. The inclusion of such fortuitous scores with scores from subjects who reacted to the stimulus words as having meaning or reference would thus confuse results from the instrument. That the testee is to react to meanings or concepts was clearly assumed in the construction of the tests, and Pressey says[8] that since the number of items checked decreases with age rather than increases, "it should be noted that the test seems to measure attitudes rather than vocabulary." This conclusion appears to be sound, but, as indicated above, this function of the test can be confused by applying the instrument to younger or duller subjects who, failing to comprehend certain of the stimuli, check them anyway. Pressey also recognizes this point in his reply to Thorndike's criticism.[9]

While attitudes and values are undoubtedly important components of emotional maturity, it seems rather premature to identify reactions

to words with as complex a construct as "emotional maturity" as it at present is described by psychologists. The authors of the Interest-Attitude tests have published evidence that combined estimates of emotional maturity (not defined) correlate from .58 to .72 with

FIG. 2



Interest Attitude scores.[8] They present descriptive data to show that "the great majority of the maladjusted cases makes scores that are either very high or very low—that is, they are unusually childish or mature for their grades." The relation of the concepts assessed by the stimulus words in the Pressey tests to intellectual maturity and to specific evidences of emotional behavior, such as characteristic mood, frequency of temper outbursts, and tendency to bicker and

fight, will need to be established in order to warrant the notions of emotional age and emotional maturity.

The foregoing paragraphs indicate that the responses of young boys probably have less value than those of adolescents, say of fourteen years and above. While younger children may make immature scores, their responses may actually be very uncritical. Indeed, the inclusion of subjects having mental ages of less than eleven or twelve, may perhaps be questioned.

SUMMARY

The present study clearly indicates that delinquent boys are more like non-delinquents who come from similar social background than delinquents are like unselected boys in general. It is probable that more careful matching of groups than was possible in this study would further reduce the differential noted by Durea. There appears to be sufficient correlation of Pressey test score with mental ability to account for some, if not all, of the delinquent's emotional retardation. Thus Durea's conclusion, that emotional retardation may be just as significant for delinquency as mental level and other factors, may have to wait for additional empirical knowledge as to what the Pressey tests test.

Probably emotional maturity, as at present defined by psychologists, is rather broad to be measured solely by word reactions, unless separate measures of specific components of emotional maturity can be shown to correlate appreciably with such word reactions. It is also suggested that the maturity reflected by reactions to the stimulus words in this test may consist in a refinement of concepts with mental maturity and experience, rather than in a differential reaction to categories of words. The usefulness of the test with young or dull subjects is questioned because of the possibility of these subjects checking verbal stimuli which they do not comprehend.

REFERENCES

1. Chambers, O. R.: "A Method of Measuring the Emotional Maturity of Children." *Ped. Sem. and Jr. of Genet. Psychol.*, Vol. XXXII, 1925, pp. 637–647.
2. Durea, M. A.: "The Emotional Maturity of Juvenile Delinquents." *Jr. Abn. and Soc. Psychol.*, Vol. XXXI, 1937, pp. 472–481.
3. Furfey, P. H.: "The Measurement of Developmental Age." *Cath. Univ. of America Educ. Res. Bull.*, Vol. II, No. 10, Dec., 1927.

4. Hilden, A. H.: *Table of Per Cent of Average Development Based on Mental Growth Units.* Minneapolis: Educational Test Bureau, 1937.
5. Kuhlmann, Fred: *Tests of Mental Development.* Minneapolis: Educational Test Bureau, 1939.
6. Owen, M. B.: "The Intelligence of Institutionalized Juvenile Delinquents." *Jr. of Juv. Res.*, Vol. XXI, 1937, pp. 199–206.
7. Prescott, D. A.: "Emotion and the Educative Process." *Amer. Council on Education*, Washington, D. C., 1938, pp. 323.
8. Pressey, S. L. and Pressey, L. C.: "Development of the Interest-Attitude Tests." *Jr. of Appl. Psychol.*, Vol. XVII, 1933, pp. 1–16.
9. Pressey, S. L.: "A Note on the Critical Note." *Jr. of Appl. Psychol.*, Vol. XXII, 1938, pp. 659–661.
10. Thorndike, E. L.: *The Teachers' Word Book.* New York: Teachers College, Columbia University, 1921, 134 pp.
11. Thorndike, R. L.: "Critical Note on the Pressey Interest-Attitude Test." *Jr. of Appl. Psychol.*, Vol. XXII, 1938, pp. 657–658.
12. Thorndike, R. L.: "Responses of a Group of Gifted Children to the Pressey Interest-Attitude Test." *Jr. of Educ. Psychol.*, Vol. XXX, 1939, pp. 587–593.
13. Weber, C. O.: "The Concept of 'Emotional Age' and Its Measurement." *Jr. of Abn. and Soc. Psychol.*, Vol. XXIV, 1930, pp. 466–471.
14. Weber, C. O.: "Further Tests of the Wells Emotional Age Scale." *Jr. Abn. and Soc. Psychol.*, Vol. XXVII, 1932, pp. 65–78.
15. Willoughby, R. E.: "A Scale of Emotional Maturity." *Jr. of Soc. Psychol.*, Vol. III, 1932, pp. 3–6.

AN ANALYSIS OF ERRORS IN WORD RECOGNITION MADE BY RETARDED READERS

ANNETTE BENNETT

School Psychologist and Reading Specialist, District #15, Long Island, N. Y.

The mental confusion which some children experience in the discrimination of printed words has been subject to many experimental investigations. Hill[1] analyzes nine of these outstanding studies and concludes that most experimenters in this field resort to a more or less subjective classification of the errors made by readers in word analysis. In some instances the terms used are subject to varying interpretations. For example, in the literature on the subject, "reversals" will be found designating all directions of mirror imaging or directional disorientation in single letters and all kinds of sequence errors. Research workers, using diagnostic tests, seldom agree absolutely on their scoring of errors as to type, since it is seldom that a particular error can be definitely classified in one error category.

More carefully controlled experimental procedures are necessary to prove scientifically the many problems involved in the analysis of the learning difficulties of children in the complex process of word discrimination. Questions such as the following are of particular interest:

What parts of words are most frequently observed by children and used as cues in word recognition?

Is the first half of the word more potent than the last half in word perception?

Have pupils who are extremely weak in word discrimination formed the habit of depending for cues upon trivial or non-characteristic details of words?

What letters and letter combinations prove most confusing?

Are words that are substituted for given words similar in form, similar in idea, or unlike in form or idea?

Is a limited perceptual span a dominant characteristic of retarded readers?

What parts of speech predominate in reading errors?

Do errors due to reversal tendencies predominate over all other wrong responses?

[1] Hill, M. B.: "Experimental Procedures in the Study of the Process of Word Discrimination in Reading." *Jour. Edu. Res.*, Vol. XXIX, No. 6, 1936, pp. 473–482.

Do median vowel errors predominate to such an extent that the hypothesis of "islands of deafness" for the short vowel sounds seems a plausible explanation?

Do errors become more "sensible" as pupils progress in reading skill?

SOURCE OF DATA IN PRESENT STUDY

The present study is an analysis of 34,274 errors made by retarded readers in word recognition. The data were collected by tutors in the field rather than from the laboratory. The niceties of a carefully controlled, scientific procedure could not, therefore, be observed, but an attempt was made to adhere to a strictly objective method in the analysis of the errors.

The children whose errors are tabulated in this report were coached individually for one period each day. Many of them were classified as non-readers who had made little or no progress in the mastery of the printed page although they averaged third or fourth grade in school placement. They were enrolled in various schools in the New York Metropolitan area.

DIAGNOSTIC TESTS

For diagnosis, oral reading tests were constructed which embodied a basic reading vocabulary. To this end the fifteen hundred words of the *Gates Primary Word List* were arranged in order of frequency of use and then grouped into units of twenty words each, with the twenty most frequently used words in Unit 1, the next twenty in Unit 2, and so on. The words in each unit were then incorporated in five or six short sentences which included all the words of that unit but made use of no words which had not previously been introduced in the graded series.

These sentences were used as preliminary diagnostic oral reading tests. The errors made in word recognition were recorded by the coach on a blank form provided for this purpose. If a pupil made five or more errors in reading the sentences based on any vocabulary unit, the coach made use of remedial reading exercises constructed for that unit and continued the remedial instruction at that level of achievement until the word difficulties had been mastered. Since lack of basic reading vocabulary was a pronounced characteristic of all of these retarded readers, the diagnosis for each pupil began at Unit 1, and proceeded at varying rates of progress to the higher units.

The findings of only the first thirty units are given in this report. These constituted five hundred ninety-four basic words (six words were

lacking, owing to clerical errors). The blank forms on which the word errors were recorded by the coaches were collected and the results summarized.

Since each pupil's progress through the reading units varied with his rate of learning, the responses of the slowest learners were decreasingly represented in the higher units. From the data available we infer that the responses of five hundred ninety-five pupils were recorded in the first ten units and that at least one hundred thirteen of these pupils progressed to the twenty-ninth or thirtieth unit.

ANALYSIS OF WORDS IN THE THIRTY VOCABULARY UNITS

Of the five hundred ninety-four words in the thirty vocabulary units, four hundred thirty-nine or seventy-four per cent were monosyllables; one hundred forty-two or twenty-four per cent were words of two syllables; seven or two per cent were words of three or four syllables. For the purposes of this study the predominance of monosyllables was a fortunate one, since the difficulties involved in syllabication were largely eliminated.

TABLE I.—WORDS OF THE THIRTY VOCABULARY UNITS CLASSIFIED AS TO PARTS OF SPEECH

Part 1			Part 2		
Analysis of words in the thirty vocabulary units			Analysis of 237 words which called forth erroneous responses		
	Number	Per cent		Number	Per cent
Nouns	216	36.3	Nouns	73	31
Pronouns	29	5	Pronouns	12	5
Adjectives	74	13	Adjectives	33	14
Adverbs	38	6	Adverbs	14	6
Verbs	199	33.5	Verbs	93	39
Preposition	29	4.8	Prepositions	9	4
Conjunctions	7	1.1	Conjunctions	3	1
Interjections	2	.3	Interjections	0	

Perhaps the absence of longer, more complex words accounted for the fact that no nonsense words were recorded among the erroneous responses. "Ready," for example, might be pronounced "reading," but such responses as "ast-rep" for "repast," or "as-tinkly" for "distinctly" are not found in the data recorded by the coaches.

In Table I, Part 1, the five hundred ninety-four words in the thirty vocabulary units are analyzed into parts of speech, as these words are commonly used. In Part 2 of the same table are analyzed two hundred thirty-seven of these words which called forth erroneous responses when they were incorporated into sentences and used in the preliminary diagnostic tests. Non-attempts were not recorded. On the whole the percentages for the different parts of speech are strikingly similar in the two parts of the table.

CLASSIFICATION OF ERRORS INTO TYPE FORMS

Classification of wrong responses on the basis of consonant errors, vowel errors, words of similar configuration, words associated in meaning, etc. was rejected in the present study since such classification gives rise to subjective interpretations and divergent conclusions. Moreover, an inspection of the responses shows that the beginnings of words, the endings of words and the presence of similar word parts of varying length are the dominant visual and auditory cues which call forth associated responses. Varying vowel sounds are ignored, as, for example, in the substitution of "what" for "that." The cluttering of syllables which is the result of a confused, unsystematic attack upon polysyllables was largely eliminated because the majority of the words were monosyllabic.

In the analysis into type errors the following procedure was followed.

(1) *Type A*.—An error which had one, two or more initial letters in common with the stimulus word but an unlike ending was tabulated under Type *A*. Examples of such words are:

"Cat" pronounced "cup."
"Brown" pronounced "blue."
"Do" pronounced "did."

(2) *Type B*.—Under the *B* Type were grouped the erroneous responses which had one or more letters at the end of the word in common with the stimulus word but which had unlike beginnings. Examples of the *B* Type are:

"Have" pronounced "love."
"That" pronounced "at."
"There" pronounced "here."

(3) *Type AB(A)*.—In the *AB (A)* type the letters at the beginning and end of the stimulus word and the erroneous response were alike, but

more like letters were found in the beginning of the words than at the end. Examples of the *AB(A)* Type are:

"Make" pronounced "made."
"Puts" pronounced "Pulls."
"May" pronounced "many."

(4) *Type AB(B).*—In the *AB(B)* Type the letters at the beginning and end of the stimulus word and erroneous response were alike, but more like letters were found at the end of the words than at the beginning. Examples of the *AB(B)* Type are:

"Goes" pronounced "gives."
"Watch" pronounced "which."
"Forest" pronounced "first."

(5) *Type AB.*—In the *AB* Type the letters at the beginning and end of the stimulus word and the erroneous response were alike both in form and number. Examples of the *AB* Type are:

"Like" pronounced "love."
"Says" pronounced "sits."
"Comes" pronounced "calls."

(6) *Median Vowels.*—Because of the emphasis which has been given to the part played by median vowels in the confusion of words it seemed desirable to separate the pure median vowel errors into a separate type rather than to group them in the *AB* Type where they logically belong. Some research workers, who stress the part played by faulty discrimination of single letter sounds, have advanced the hypothesis that "islands of deafness" for the short vowel sounds account for the many errors made in words which differ in median vowels only. The sixth, or Median Vowel, Type was included in this tabulation to see if such errors do predominate over the other types. In this classification a final "*s*" error or "*ed*" error was disregarded if the erroneous response and the stimulus word were otherwise of the pure median vowel type. Examples of Type 6 are:

"Cup" pronounced "cap."
"Runs" pronounced "ran."
"Lives" pronounced "loves."
"Lifted" pronounced "left."

(7) *Reversals.*—Reversal errors were classified under two headings.
(*a*) Those due to reversals of the initial consonants *b*, *d*, *p*, *q*.

(*b*) Errors due to the reading of the whole word or word part in the right to left sequence. Examples of Type 7 are:

"Big" pronounced "dog."
"Made" pronounced "am."
"For" pronounced "of."

(8) *Final "s" Errors.*—In Type 8 were grouped the final "*s*" errors, where the omission or the addition of final "*s*" was the sole error, as,

"News" pronounced "new."
"Eye" pronounced "Eyes."

(9) *Substitutions.*—The first eight types absorbed many responses which might otherwise have been classified as substitutions. In fact the first eight types included the erroneous responses on all but the following ten words:

the, cat, a, out, my, no, or, me, his, of.

The responses to these words were classified under substitutions, since the erroneous response had neither the beginning nor the ending of the word in common with the stimulus word. Some of them had at least one striking letter form in common with the stimulus. Examples of Type 9 are:

"The" pronounced "a."
"Of" pronounced "from."
"His" pronounced "the."

When the 34,274 errors were tabulated under the nine types, the distribution was found to be as follows:

TABLE II.—TYPE ERRORS ANALYZED

Type	Number of errors	Per cent of total
(1) *A*	10,572	31
(2) *B*	5,478	16
(3) *AB(A)*	2,147	6
(4) *AB(B)*	1,630	4
(5) *AB*	2,521	8
(6) Median vowels	5,041	15
(7) Reversals	3,815	12
(8) Final "*s*"	1,869	5
(9) Substitutions	1,201	3

In the *A* Type the stimulus word might have one to five initial letters in common with the erroneous response. This type included thirty-one per cent of the total number of errors and, therefore, predominated as a single type form.

The *B* Type, in which one to six letters at the end of the erroneous response were like those at the end of the stimulus word, constituted sixteen per cent of the total number of errors. A comparison of the total errors in the *A* and *B* types shows that there is almost a two to one chance that the beginning of the word will be more dominant as a cue in word recognition than the ending.

Of the 5478 errors included under the *B* Type it was found that 1635 of these were due to confusion of single initial consonant or initial vowel sounds, while 2299 errors of the *B* Type were due to confusions in which the initial consonant blends were involved. The total of 3934, or 71.8 per cent of the *B* errors, could have been eliminated if there had been no confusion of initial letter sounds.

The *B* errors in which the erroneous response differed from the stimulus word in initial single letter only, are given in the following list: (The differences in vowel sounds, as in "goes" pronounced "does," are ignored in this tabulation.)

B Errors Due to Confusion of Initial Single Letters

Stimulus	Response	Frequency	Stimulus	Response	Frequency
Has	was	61	Cut	put	55
Cut	out	35	Right	night	77
On	in	95	Lives	gives	42
Gives	lives	37	At	it	61
Goes	does	26	Night	right	40
Our	your	49	We	he	95
Does	goes	115	Took	look	155
Look	took	28	Save	have	32
How	now	191	Every	very	128
Very	every	132	Your	our	28
Donkeys	monkeys	96	Now	how	19
Watch	catch	21	Cent	kent	17

If to the 3934 errors of the *B* Type, due to the confusion of initial letter sounds, are added 651 similar errors of the *AB*(*B*) Type, as, for example, "grate" for "gate" and 311 reversals of initial *b* and *p* ("big" for "pig"), a total of 4896 errors, constituting fourteen per

cent of the total tabulation, were due solely to confusion of the initial letters of words.

A comparison of the total errors in the *AB*(*A*) Type and the total errors in the *AB*(*B*) Type (Table II) shows a ratio of about three to two in favor of the longer word part at the beginning of the word *versus* the ending as a dominant cue in word recognition.

When the third, fourth, fifth and sixth types are added (Table II), it will be seen that the errors in which the stimulus word and the response have like beginning and ending constitute thirty-three per cent of the 34,274 errors. The sixth, or Median Vowel Type, accounts for forty-five per cent of these combined groups, or fifteen per cent of the total tabulation. Evidently sharpness in visual discrimination which recognizes likenesses and differences in the middle as well as at the beginning or end of words is a product of growth. This study shows that poor word readers are particularly deficient in this respect.

Reversal errors proved to be a close rival of median vowel errors and constituted twelve per cent of the total tabulation. These errors were divided into: (1) Reversals of the initial consonants *b*, *d*, *p*, *q* and (2) errors due to the reading of the whole word or word part in the right to left sequence. In the letter reversals, "*d*" was substituted for "*b*" seven hundred sixty-eight times; "*b*" for "*d*" six hundred fifteen times; and "*b*" for "*p*" three hundred eleven times. These letter errors were most pronounced in the first four units and seemed to be eliminated after the tenth vocabulary unit was reached. There was a slight preponderance of reversals (2) over letter reversals (1). Because of the frequency with which such words as was-saw, on-no are used, and because of the ease with which such words can be confused, one for the other, it is not surprising to find that word reversals persisted up to the thirtieth unit in this study. At this stage the sequence from left to right in word recognition seemed to be fairly well established, but the right to left sequence in the recognition of these particular words, when once acquired in the learning process, showed a tendency to persist and to erupt unexpectedly, even under the most careful remedial instruction.

Substitutions, as explained above, accounted for only three per cent of the errors and the omission or addition of final "*s*" made up the remaining five per cent.

To note any striking trend in the sub-totals, the frequency of errors under each type was converted into percentage of errors for each vocabulary unit. These results are given in Table III.

TABLE III.—PERCENTAGE OF ERRORS FOR EACH VOCABULARY UNIT

Type	*A*	*B*	*AB(A)*	*AB(B)*	*AB*	Medium V	Rever	Final *S*	Substi-tutions	Fre-quency
Unit										
1	42	12	0	0	0	12	23	0	11	1,982
2	25	11	14	0	11	10	13	9	7	3,790
3	13	7	1	1	12	22	22	15	7	2,555
4	21	37	5	5	3	2	22	3	2	4,104
5	28	37	3	5	9½	0	9½	4	4	1,986
6	28	3	14	0	28	4	10	9	4	2,176
7	46	22	0	14	0	0	3	12	3	1,058
8	44	0	0	0	9	44	2	1	0	2,570
9	16	16	0	0	8	42	14	1	3	1,759
10	52	2	11	1	0	4	20	3	7	2,599
11	51	15	28	0	3	3	0	0	0	1,206
12	30	7	0	14	0	47	0	0	2	855
13	23	36	4	14	0	0	4	19	0	790
14	21	44	0	10	4	7	0	14	0	882
15	45	19	0	4	0	19	0	11	2	1,079
16	14	14	19	0	0	53	0	0	0	418
17	0	55	14	0	16	15	0	0	0	693
18	45	5	17	5	23	0	0	5	0	414
19	19	0	7	0	0	55	0	19	0	305
20	10	23	42	15	10	0	0	0	0	561
21	12	0	21	7	13	28	19	0	0	253
22	46	8	0	21	12	13	0	0	0	224
23	36	0	5	23	0	36	0	0	0	542
24	17	0	5	73	5	0	0	0	0	475
25	84	0	0	0	0	16	0	0	0	325
26	15	17	0	22	20	18	0	8	0	248
27	50	0	0	0	0	50	0	0	0	268
28	0	0	0	0	100	0	0	0	0	19
29	79	0	0	21	0	0	0	0	0	191
30	54	12	0	0	0	23	11	0	0	147
										34,274

Table III shows that few trends in the learning process are outstanding except that reversal errors cease to be dominant after the tenth unit; substitutions fall off to a negligible number about the same time and the final "*s*" error seems to be largely corrected after the nineteenth unit. What is most marked is the disruptive effect of certain words as they are encountered in the successive units. For example, the "began-begin" stimulus and response in the sixteenth unit and "begin-began" in the nineteenth unit pile up the scores for median vowel errors at these points. The use of a familiar word "playing" for "planting" caused all the errors in the twenty-eighth unit. If any trend is at all consistent, it is the gradual widening of the perception span for the initial letters of words. That is to say, in

Unit	Basic word	Erroneous response	Frequency	Type error
1	runs	ran	204	*M*
	big	dog	226	*R* (letter)
	the	this	208	*A*
2	of	for	251	*R*
	made	make	384	*AB*(*A*)
	love	live	302	*M*
	says	said	341	*A*
3	lives	loves	254	*M*
	story	store	233	*A*
4	for	of	457	*R*
	here	there	544	*B*
	there	here	455	*B*
	doll	ball	470	*R*
5	no	on	114	*R*
	says	say	83	Final "*s*"
	what	that	144	*B*
6	pulls	puts	275	*AB*(*A*)
	saw	was	232	*R*
	horse	house	255	*AB*
7	dolls	doll	128	Final *S*
	does	goes	115	*B*
8	came	come	505	*M*
	may	my	184	*AB*
	lived	lives	235	*A*
9	rode	ride	341	*M*
	went	want	223	*M*
10	lay	laid	595	*A*
	pig	big	311	*R*
	chicken	children	263	*AB*(*A*)
11	their	they	296	*A*
	when	then	146	*B*
12	Street	store	145	*A*
	want	went	250	*M*
13	rabbits	rabbit	114	Final *S*
	then	when	171	*B*
14	how	now	191	*B*
	every	very	128	*B*
15	falls	fell	180	*M*
	soon	some	139	*A*
16	began	begin	111	*M*
17	ate	eats	148	*B* (?) or Sub.
18	wakes	walks	69	*AB*(*A*)
19	begin	began	146	*M*
20	stayed	stand	105	*AB*(*A*)
21	washed	wished	71	*M*
22	fast	first	46	*AB*(*B*)
23	hard	heard	130	*AB*(*B*)
24	tied	tried	142	*AB*(*B*)
25	safe	safely	59	*A*
26	sheep	sleep	55	*AB*(*B*)
27	blew	blow	67	*M*
	blew	blows	67	*M*
28	planting	playing	19	*AB*
29	cool	cold	113	*A*
30	birdie	bird	54	*A*

unit 1, many children respond with a word which may have one letter only in common with the stimulus word, as, for example,

"Brown" pronounced "big."
"Cup" pronounced "cat."

In the last five units no errors are found in which the stimulus word and the response have not two, three, four or more initial letters in common.

The list on page 34 gives the words in the successive units which had the maximum erroneous responses at each level: (Two or more words are given for each unit up to the fifteenth unit and one word for each of the subsequent.)

These fifty-four words and their responses, when classified as to type, are found to be distributed as follows:

	I	II	III	IV	V	VI	VII	VIII	IX
Type.....	*A*	*B*	*AB(A)*	*AB(B)*	*AB*	Median	Reversals	Final *S*	Substitutions
Number ...	12	9	5	4	3	11	8	2	0
Per cent...	22	17	9	7	6	20	15	4	0

These results should be compared with the distribution given in Table II. It will be seen that the *A* Type loses nine per cent; the *B* Type gains one per cent; the *AB(A)* Type gains three per cent; the *AB(B)* type gains three per cent; the *AB* Type loses two per cent; the Median Vowel Type gains five per cent; Reversals gain three per cent; Final *S* loses one per cent; Substitutions lose three per cent.

The two distributions are, on the whole, fairly similar.

Although the pupils who took part in this research project were widely scattered in many schools in the New York Metropolitan area, it is amazing to see with what uniformity a stimulus word called forth the same erroneous response on the part of many pupils. This may be due to the fact that these basic words had frequently been encountered by the pupils in primer material which they had tried, but failed to master. Basic words which had any word part in common with the stimulus word easily became confused and, as a result, wrong habits of reacting to these cues had been established.

In the data which were presented in this study, no record was kept of the words in the vocabulary units which could not be pronounced

at all. We may infer that there were quite a number of these, but the inclusion of these non-pronunciations would have thrown no light on the errors which the children were making in word perception.

Only eighty-two of the two hundred thirty-seven words elicited two or more different responses as is shown in the following tabulation:

1. Number of words eliciting one erroneous response..... 155
2. Number of words eliciting two erroneous responses..... 50
3. Number of words eliciting three erroneous responses..... 21
4. Number of words eliciting four erroneous responses..... 8
5. Number of words eliciting five erroneous responses..... 3

The eleven words which comprise the fourth and fifth groups are all found in the first six vocabulary units and are, therefore, among the one hundred twenty words most frequently encountered by children in primer and first-grade material. These eleven words are:

cut says there which gets gives on
out no that goes

Although dominant letters and word parts play an important part in word recognition, nevertheless, the structure of the context in which the word is incorporated seems also to operate in some intangible way to govern the response, so that a verb stimulus, for example, calls forth in most cases a verb in response, even though it is erroneous. Occasionally a word which is closely associated in thought with the stimulus word may be given, even if it is of a different part of speech, as in "grass-green." Especially is this true if a dominant word part, such as "gr," is common to the two words.

A subjective interpretation is necessarily involved in any attempt to determine which responses were relevant or irrelevant to the ideational context in which they were incorporated. The writer is of the opinion that forty-one per cent of the erroneous responses were closely associated in thought with the stimuli. Of these relevant responses only a very negligible number differed from the stimuli in part of speech. About fifty per cent of the irrelevant responses were also of the same part of speech as the stimuli.

Teachers will recognize that the responses given in these tabulations are the ones which they repeatedly hear given by retarded readers when these basic words are encountered. They also know that the readers are often able to give the correct response when their attention is called to an error.

Experimental data and experience, therefore, seem to justify the conclusion that the tendency not to inhibit an associated response until

a word is clearly perceived as a whole is an important factor in accounting for retardation in word recognition and in the resulting mental confusion in reading.

SUMMARY

An analysis is presented of 34,274 errors made by retarded readers in the recognition and pronunciation of two hundred thirty-seven basic words in context. Classification of erroneous verbal responses on the basis of consonant errors, vowel errors, words of similar configuration etc. was rejected, since it was found that the beginning of words, the ending of words, and the presence of similar word parts of varying length are the dominant visual and auditory cues which call forth associated responses. Total non-pronunciations were not included in this study.

Errors were classified into nine types, namely:

(1) *A*.—Errors which had one, two, or more initial letters in common with the stimulus word but unlike ending.

(2) *B*.—Errors which had one or more letters at the end of the word in common with the stimulus word but unlike beginning.

(3) *AB*(*A*).—Letters at the beginning and end of the stimulus word and response were alike, but more like letters were found in the beginning of the words than at the end.

(4) *AB*(*B*).—Letters at the beginning and end of the stimulus word and erroneous response were alike but more like letters were found at the end of the words than at the beginning.

(5) *AB*.—Letters at the beginning and end of the stimulus word and erroneous response were alike both in form and number.

(6) *Median Vowel Type*.—Stimulus and erroneous response differed in median vowel only.

(7) *Reversals*.—Classified under two headings: (a) reversal of initial consonants; (b) reading of whole word or word part in the right to left sequence.

(8) *Final S*.—Omission or addition of final "*s*" was sole error.

(9) *Substitutions*.—Erroneous response had neither the beginning nor the ending of the word in common with the stimulus word.

The *A* type included thirty-one per cent of the total number of erroneous responses: The *B* Type, sixteen per cent; the *AB*(*A*) Type, six per cent; the *AB*(*B*) Type, four per cent; the *AB* Type, eight per cent; Median Vowel Type, fifteen per cent; Reversals, twelve per cent; Final "*s*," five per cent; Substitutions, three per cent.

The data indicate that the beginning and ending of words are most frequently used as cues in word recognition and that there is almost a two to one chance that the beginning of the word will be more potent as a dominant cue than the ending.

Errors in which the stimulus word and the response have like beginning and ending constitute thirty-three per cent of the total tabulation. Forty five per cent of this group, or fifteen per cent of the total tabulation, fall into the type classified as Median Vowels. Failure to observe the middle section of words is, therefore, an important factor in mispronunciations, especially of monosyllables. "Islands of Deafness" for the short vowel sounds may account, in individual cases, for some errors of this type.

Reversals proved to be a close rival to median vowel errors and constituted twelve per cent of the total tabulation. Reversal of words, such as saw-was, no-on, proved to be more frequent and more difficult to eliminate than letter reversals.

As the pupils developed skill in word discrimination a few trends could be discerned in the learning process, such as the following: (1) reversal errors ceased to be dominant after the tenth vocabulary unit; (2) final "*s*" errors seemed to be largely corrected after the nineteenth unit; (3) the perception span for the initial letters of words gradually widened so that single letters in the stimulus word ceased to be the sole cue for the erroneous response as was true in the first units.

Although dominant letters or word parts played a very important part as visual and auditory cues in word recognition and pronunciation, nevertheless, the structure of the context in which the stimulus word was incorporated played its rôle in governing the verbal response. The writer estimates that forty-one per cent of the erroneous responses were closely associated in thought with the stimuli and that, of the irrelevant responses, fifty per cent were of the same part of speech as the stimuli. Of the relevant responses, only a very negligible number differed from the stimuli in part of speech.

Thus it is seen that word recognition and pronunciation are complex mental processes in which the neural centers involved in language usage play a controlling part. Errors do not occur in an haphazard way, but are governed by the context in which the stimuli are incorporated, and by unfortunate learning habits which the pupil has developed in the process of reading. A pronounced characteristic of pupils retarded in word recognition seems to be the tendency not to inhibit associated responses until a word is clearly seen in all its parts—beginning, middle and ending.

TERMS THAT ARE CONSIDERED IMPORTANT FOR BEGINNING STUDENTS OF PSYCHOLOGY[1]

G. R. THORNTON AND J. S. THORNTON
Purdue University

A problem that frequently bothers teachers is the problem of determining what is worth teaching. Since the teacher usually decides upon areas of subject-matter for which the student is to be held responsible and upon the mastery of which the student is to be judged, it is only fair that the decision concerning precisely for what the student is to be held responsible should be reached with some care. This decision is especially difficult for teachers of elementary psychology. There are varying conceptions of the function of the first course in psychology, and there is so much that could be included in it. The survey, results from which are reported in this paper, was undertaken in the hope that there might emerge a list of fifty or a hundred terms that most teachers of general psychology would agree should be stressed in the beginning course.[2] This hope was only partially realized. Instead of clearcut agreement, there was forthcoming direct and indirect evidence of difference of opinion concerning the terms and topics that should be emphasized in the first course.

METHOD OF CONDUCTING THE SURVEY

The survey was conducted during the Winter of 1940–1941. Eight then current textbooks for the first course in psychology were selected as representative of the textbook material being used in American colleges and universities. The eight books selected are listed in the bibliography at the end of this article. These textbooks were scanned with the purpose of picking out from them all the terms used that we believed might be unknown to an average beginning student. No claim is made for the complete accuracy of the original tabulation.[3]

[1] Thanks are due to the thirty-nine psychologists who generously coöperated in the survey.

[2] It is not assumed that this survey is equivalent to a survey of the subject-matter that should be covered. As suggested by some respondents in the present survey, a survey of topics or subject-matter areas that should be included in the first course might be useful.

[3] Arbitrary decisions were necessary in determining whether similar terms should be counted as the same term or as different terms. *Neurotic* and *neurosis*, *e.g.*, were counted as the same term, but *thyroid* and *thyroxin* were considered different terms. Undoubtedly, we have been guilty of several minor inconsistencies; *e.g.*, we have listed *pituitary gland* and *thyroid*.

Undoubtedly there were some terms that were used in one or two of the textbooks that we overlooked. There probably were a few terms that occurred in three or more books that were overlooked.

Parenthetically, it may be remarked that considerable variation was noted from book to book in the terms that were introduced. Of twelve hundred forty-six terms each of which was found in at least one book, eight hundred eighty-one were found in only one or two of the eight textbooks. Three hundred sixty-six terms appeared in three or more books. Of these latter, relatively few were found in all eight textbooks.

From the tabulation of terms found in the various books, we composed a list of the terms that appeared in at least three of the eight textbooks. Of these terms some were technical, and some were not. Some were given considerable attention, and others were used only incidentally. This list of terms, arranged in alphabetical order, was sent to seventy-one teachers of general psychology with the following request:

> Please go through the list and check the terms which you think should be emphasized in the beginning course. Check approximately one hundred terms which you consider most important. Omit from the ones you check (1) those that should be postponed until later courses, (2) those that students will already understand adequately, (3) those that are of little importance for a beginning student in psychology.

The list was sent to one psychologist at each of seventy-one different universities or colleges for which four or more persons were listed in the 1940 *A.P.A. Directory*. In deciding the person at each school to whom the list should be sent, we considered the following criteria, weighting each in importance approximately in the order listed (the first criterion receiving greatest weight): The person should (1) be listed in the *A.P.A. Directory*, (2) teach general psychology, (3) have a rank of associate professor or higher, (4) be a full member of A.P.A., (5) have received his doctoral degree since 1925. In many instances, of course, no person could be found who satisfied all of these criteria.

Of the seventy-one persons to whom the list was sent, thirty-nine responded. These thirty-nine persons[1] are, with one exception, teachers of general psychology in colleges and universities scattered throughout the United States. Classified according to sections of the

[1] The following statements are based on the 31 of the 39 who could be definitely identified.

country, eleven are located in the East, two in the South, thirteen in the Mid-West, and five in the West. About two-thirds of them are members of A.P.A., and the rest are associate members. About half of them have the rank of professor, and most of the rest have the rank of associate professor. Most of them received their doctoral degree since 1920, about half of them since January, 1930.

RESULTS

The persons who checked the list did not succeed in checking exactly one hundred terms: Fourteen persons checked between seventy-six and a hundred words; twelve persons checked between one hundred one and one hundred twenty-five; three persons checked between one hundred twenty-six and one hundred fifty; three persons checked between one hundred fifty-one and one hundred seventy-five; two persons checked between one hundred seventy-six and two hundred; and five persons checked between two hundred one and three hundred words. The particular words checked were tabulated according to the groupings indicated above; but there were no evident trends that were different from the trends for the total group. In summarizing the results, therefore, we list only the tabulation for the total group of respondents.

Since readers will differ in the particular terms or groups of terms in which they are interested, we present the list in alphabetical order in order to facilitate the location of particular items. For each term there are indicated (1) the number of persons who checked the term as important, (2) the rank order of the term for number of checks received, and (3) the number of books, of the eight investigated, in which the term is listed in the index.[1] (We assume that whether a term is listed in the index or not is some indication of the degree of importance attached to the term by the author of the text.)

The reader should remember that a term may have failed to be checked for any one of three reasons: (1) It was considered relatively unimportant. (2) It was believed the term should be postponed until a later course. (3) It was believed that students would already understand the term sufficiently well. The number of times each term was checked indicates the number of teachers, of the thirty-nine

[1] A term was considered listed in the index only when it appeared in the alphabetical position of one of the important words in the term. Close derivatives of the same item (*e.g.*, *auditory* and *audition*) were treated as identical. Many terms were found in each book which did not appear in the index.

replying, who thought the term should be given special emphasis in the beginning course in psychology.

One limiting feature of the results that should be borne in mind is that in some cases closely related terms occurred in the list. For example, *conditioning* and *conditioned response* occurred at different points in the list; likewise, both *Binet test* and *Stanford-Binet test* were listed; again, *sense organ*, *receptor*, *exteroceptor*, *interoceptor*, and *proprioceptor* were all listed separately. What effect such overlapping had upon the results is unknown. No directions were given concerning such cases. Persons who considered two terms synonymous may have checked both terms or they may have chosen one of the terms. In the latter case the alphabetical order of the check list may have influenced the choice.

Besides the vote for particular terms, readers may be interested to note the following: (1) What sort of terms received most checks and what sort of terms received fewest checks; (2) the number of checks received by (*a*) physiological and biological terms, (*b*) statistical terms, (*c*) social-psychological terms, (*d*) terms related to methodology in psychology; (3) evidences of changes in emphases which have occurred in the past few years.

Space does not permit extended discussion of these points. We shall limit ourselves to a few observations concerning point (2). Of the three hundred sixty-six terms in the list, we have classified twenty terms as referring to methodology, fourteen as statistical terms, thirty as social-psychological terms, and eighty-nine as physiological or biological terms. These figures suggest the fact that physiological and biological terms are used frequently in beginning texts in psychology. The apparent preponderance of such terms may reflect in part the method of selection used in this study—terms from other areas may have been omitted more frequently because they seemed more likely to be familiar to beginning students. Of these physiological and biological terms, however, only six did not appear in any index, and seventy-one of them appeared in three or more indexes of the eight textbooks examined. The mean number of indexes in which physiological and biological terms appeared was 3.9, while the mean number of indexes for statistical, methodological, and social-psychological terms was 3.2, 3.4, and 4.5, respectively.

When checked by the respondents, the physiological and biological terms tended to receive fewer than their proportionate share of checks. Approximately two-thirds of them received less than the median rank

for number of checks for all terms; forty-five of them received nine or less checks; twenty of them received four or less checks. In contrast to this, social-psychological, statistical, and methodological terms tended to receive more than their proportionate number of checks. The mean number of checks received by the respective groups of terms was as follows: Physiological and biological terms, 10.9; social-psychological terms, 15.4; methodological terms, 16.4; and statistical terms, 18.4.

These results suggest that, in contrast to the emphases given by textbook writers, the respondents tended to emphasize statistical, methodological, and social-psychological terms more and physiological and biological terms less. Since the textbooks were written before the reactions were obtained from the respondents, these results perhaps indicate a trend away from emphasis upon physiological and biological concepts and a trend toward emphasis upon statistical, methodological, and social-psychological concepts.[1]

On pages 44–48 are the alphabetical list of the terms which appear in at least three of eight current textbooks in elementary general psychology, and for each term the number of psychologists (of thirty-nine respondents) who believe the term should be emphasized in the first course ("Checks"), the rank order of each term for number of checks received, and the number of books in which each term is listed in the index ("Index").

When the project reported in this paper was begun, we hoped that the results, by indicating a limited number of terms that are of outstanding importance, might prove to be one step toward the ultimate goal of a normative dictionary of important terms in general psychology. Although that hope has not been fulfilled, it is still our belief that such a normative dictionary would be a boone to instructors and students in elementary psychology and through them might lead in time to greater uniformity of usage among psychologists generally. By a "normative dictionary" we mean a dictionary giving for each term the one meaning that most psychologists are agreed should be attached to the term. Such a dictionary would of necessity be the product of the coöperative effort of many psychologists, and probably could best be achieved through some committee of the American Psychological Association. In order that such a dictionary should

---

[1] The Pearson coefficient of correlation between number of checks given by respondents and number of indexes in which a term appeared, is .56.

Term	Checks[1]	Rank order	Index	Term	Checks[1]	Rank order	Index
Ability	16	138.5	3	Average	13	192	2
Abnormal	23	53.5	3	Axon	14	174.5	4
Abstraction	11	226.5	3	Basilar membrane	11	226.5	5
Accommodation	12	205.5	4	Beats	5	318	2
Achievement	12	205.5	2	Behaviorism	14	174.5	4
Achievement tests	22	60.5	3	Bimodal	13	192	1
Achromatic	7	292.5	2	Binet test	24	44	5
Acromegaly	2	356.5	5	Binocular	13	192	2
Activity	6	305	4	Blind spot	11	226.5	3
Adaptation	31	6.5	6	Brain	10	244.5	7
Adjustment	30	9.5	2	Brightness	8	280	5
Adolescence	6	305	4	Central fissure	3	347.5	0
Adrenal cortex	5	318	3	Central nervous system	19	95.5	0
Adrenalin	18	112.5	4	Cerebellum	12	205.5	6
Adrenals	16	138.5	8	Cerebral cortex	22	60.5	4
Adrenin	9	263.5	6	Cerebral hemisphere	8	280	1
Aerial perspective	7	292.5	3	Cerebrum	14	174.5	5
Affective	21	69.5	3	Character	12	205.5	5
Afferent nerve	18	112.5	3	Chroma	9	263.5	1
After-image	24	44	6	Chromatic	5	318	2
Aggressive	12	205.5	2	Chromosomes	8	280	4
All or none law	15	153.5	3	Chronological age (CA)	14	174.5	2
Ambiguous figures	14	174.5	4	Cochlea	14	174.5	5
Ambivert	11	226.5	4	Coefficient of correlation	29	15	4
Amnesia	23	53.5	6	Color blindness	24	44	7
Amplitude	4	332	0	Color constancy	19	95.5	3
Analysis	6	305	2	Color wheel	6	305	3
Aphasia	13	192	6	Compensation	24	44	8
Aptitude test	24	44	4	Complementary	11	226.5	2
Aqueous humor	3	347.5	1	Compulsions	10	244.5	3
Army alpha examination	17	127	7	Concept	15	153.5	6
Army beta examination	7	292.5	4	Conditioned response	33	3.5	7
Ascendance	10	244.5	6	Conditioning	29	15	4
Association	26	30	7	Cones	10	244.5	4
Attention	26	30	8	Conflicts	27	24	6
Attitudes	29	15	7	Conscious	14	174.5	6
Auditory	11	226.5	4	Constancy of IQ	24	44	6
Auditory center	4	332	1				
Autonomic nervous system	28	28.5	6				

[1] The range of checks is from 0 to 34; the median term received 14 checks.

Term	Checks[1]	Rank order	Index	Term	Checks[1]	Rank order	Index
Control group	27	24	2	Experience	12	205.5	6
Controlled association	16	138.5	4	Experiment	16	138.5	3
Convergence	11	226.5	5	Experimental method	27	24	5
Cornea	4	332	4	Exteroceptor	18	112.5	4
Correlation	21	69.5	7	Extroversion	22	60.5	4
Cortex	7	292.5	4	Extrovert	17	127	3
Cortical	8	280	0	Fantasy	18	112.5	6
Cortin	4	332	4	Fatigue	11	226.5	6
Creative thinking	19	95.5	2	Feebleminded	25	35	7
Cretinism	14	174.5	6	Feeling	12	205.5	3
Criterion	10	244.5	2	Figure and ground	22	60.5	4
Dark-adaptation	6	305	4	Fixation	10	244.5	2
Data	5	318	2	Forgetting	19	95.5	8
Day-dreaming	15	153.5	3	Fovea	8	280	4
Delayed conditioned response	6	305	3	Fraternal twins	10	244.5	2
Delayed reaction experiment	17	127	5	Free association	23	53.5	6
Delusion	18	112.5	6	Frequency	10	244.5	6
Dementia praecox	11	226.5	4	Frontal lobe	8	280	3
Dendrites	11	226.5	3	Frustration	27	24	4
Dependent variable	11	226.5	2	Fundamental	3	347.5	2
Depression	8	280	2	Fusion	7	292.5	0
Diagnostic	4	332	0	Galvanometer	6	305	1
Differential threshold	12	205.5	3	Generalization	18	112.5	5
Differentiate	6	305	2	Genes	8	280	5
Dimension	7	292.5	2	Genetic method	16	138.5	3
Discriminate	20	81.5	6	Genius	16	138.5	3
Distraction	12	205.5	6	Gestalt	24	44	6
Dream	15	153.5	7	Goal	17	127	4
Effector	20	81.5	6	Gonads	7	292.5	6
Efferent nerve	14	174.5	3	Group test	12	205.5	4
Effort	1	361	2	Habit	23	53.5	6
Eidetic image	12	205.5	6	Hallucination	18	112.5	6
Emotion	24	44	8	Halo effect	14	174.5	4
Emotional response	21	69.5	3	Harmony	1	361	1
Empathy	12	205.5	5	Heredity	20	81.5	6
Endocrine glands	27	24	5	Hormone	15	153.5	3
Environment	20	81.5	7	Hue	15	153.5	5
				Hyperthyroid	9	263.5	2
				Hypnosis	14	174.5	7
				Hypoactive	3	347.5	0

[1] The range of checks is from 0 to 34; the median term received 14 checks.

Term	Checks[1]	Rank order	Index	Term	Checks[1]	Rank order	Index
Hypothesis	16	138.5	5	Light adaptation	6	305	3
Hypothyroid	9	263.5	3	Limen	11	226.5	3
Identical twins	18	112.5	5	Localization	16	138.5	5
Identification	16	138.5	3	Locomotor ataxia	1	361	2
Idiot	18	112.5	4	Maladjustment	20	81.5	4
Illusions	30	9.5	8	Manic-depressive	10	244.5	3
Image	13	192	5	Maternal instinct	2	356.5	4
Imagery	15	153.5	4	Maturation	33	3.5	8
Imagination	15	153.5	5	Maze	19	95.5	6
Imbecile	18	112.5	3	Mean	21	69.5	1
Incentive	19	95.5	5	Median	18	112.5	2
Independent variable	10	244.5	2	Medulla	4	332	4
				Melody	3	347.5	2
Individual differences	22	60.5	8	Memorizing	14	174.5	2
				Memory	19	95.5	8
Inferiority complex	20	81.5	5	Memory image	10	244.5	2
Inhibition	27	24	7	Mental activity	5	318	3
Innate	19	95.5	0	Mental age (MA)	34	1	5
Insight	29	15	6	Metabolism	8	280	3
Instinct	21	69.5	8	Mind	10	244.5	4
Insulin	3	347.5	3	Minor premise	0	364	1
Integration	15	153.5	4	Monocular	6	305	0
Intelligence	26	30	8	Mood	13	192	5
				Mood cycle	2	356.5	1
Intelligence quotient (IQ)	30	9.5	8	Moron	20	81.5	4
Intercorrelation	8	280	1	Motivation	29	15	8
Interests	11	226.5	7	Motive	17	127	5
Interoceptor	13	192	4	Motor learning	16	138.5	3
Interview	3	347.5	2	Motor nerve	7	292.5	2
Introspection	19	95.5	6	Motor neurone	6	305	1
Introversion	23	53.5	5	Mueller-lyer illusion	4	332	4
Introvert	10	244.5	3	Myxedema	4	332	3
Intuition	1	361	3	Negative adaptation	15	153.5	3
Iris	3	347.5	4				
James-Lange theory	14	174.5	6				
Kinesthetic	27	24	4	Negative afterimage	17	127	3
Language	10	244.5	8	Negative transfer	14	174.5	3
Law of effect	24	44	6	Nerve impulse	20	81.5	4
Learning	24	44	8	Nervousness	3	347.5	1
Learning curve	30	9.5	6	Nervous system	11	226.5	8
Lens	3	347.5	4	Neurone	19	95.5	6
Lie detector	10	244.5	7				

[1] The range of checks is from 0 to 34; the median term received 14 checks.

Term	Checks[1]	Rank order	Index	Term	Checks[1]	Rank order	Index
Neurotic	20	81.5	4	Plateau	21	69.5	5
Nonsense syllable	17	127	2	Prenatal	10	244.5	1
Normal distribution	33	3.5	5	Primary hues	11	226.5	4
Nuclei	1	361	0	Projection	17	127	5
Number ability	4	332	2	Proprioceptor	19	95.5	6
Objective	16	138.5	1	Psychiatry	21	69.5	3
Observation	11	226.5	5	Psychoanalysis	23	53.5	6
Obstruction method	12	205.5	4	Psychogalvanic response (galvanic skin response)	20	81.5	7
Olfactory	9	263.5	4				
Optic nerve	6	305	2				
Organism	15	153.5	4	Psychograph	12	205.5	2
Overlearning	18	112.5	5	Psychology	21	69.5	7
Overtone	9	263.5	4	Psychotic	14	174.5	2
Pain	9	263.5	7	Pupil	4	332	2
Paired comparisons	9	263.5	4	Pursuit movement	4	332	1
Paradoxical cold	6	305	4	Questionnaire	14	174.5	4
Parallax	5	318	3	Rank order	11	226.5	1
Paranoid	12	205.5	3	Rapport	8	280	0
Parasympathetic nervous system	9	263.5	2	Rating scale	18	112.5	5
				Rationalization	28	28.5	7
Parathyroid gland	3	347.5	6	Reaction	12	205.5	5
Pathological	7	292.5	0	Reaction time	25	35	7
Perception	33	3.5	8	Reasoning	31	6.5	8
Performance tests	20	81.5	6	Recall	26	30	7
Peripheral nervous system	9	263.5	0	Receptor	24	44	7
				Recognition	22	60.5	6
Periphery	5	318	2	Reflex	18	112.5	8
Perseveration	9	263.5	5	Reflex arc	14	174.5	6
Personality	29	15	8	Refractory period	14	174.5	3
Personality trait	25	35	8	Regression	14	174.5	4
Phase	2	356.5	2	Reliability	24	44	4
Phenomenon	8	280	0	Repression	21	69.5	7
Phi phenomenon	12	205.5	4	Response	14	174.5	6
Phobia	15	153.5	3	Retention	21	69.5	7
Phrenology	14	174.5	6	Retina	15	153.5	6
Physiognomy	11	226.5	4	Retinal disparity	9	263.5	2
Physiological	8	280	5	Retinal rivalry	5	318	1
Physiological limit	18	112.5	3	Retroactive inhibition	22	60.5	6
Physiological zero	4	332	3				
Physique	9	263.5	5	Rhythm	3	347.5	4
Pitch	13	192	6	Rods	9	263.5	4
Pituitary gland	11	226.5	8	Saccadic movement	5	318	2

[1] The range of checks is from 0 to 34; the median term received 14 checks.

Term	Checks[1]	Rank order	Index	Term	Checks[1]	Rank order	Index
Saturation	8	280	6	Suggestibility	14	174.5	1
Science	15	153.5	3	Suggestion	17	127	7
Sclerotic coat	3	347.5	3	Syllogism	4	332	4
Semicircular canals	15	153.5	5	Symbol	16	138.5	4
Sensation	18	112.5	5	Sympathetic nervous system	18	112.5	4
Sense organ	18	112.5	3	Synapse	19	95.5	5
Sensitivity	11	226.5	4	Synesthesia	12	205.5	4
Sensory	9	263.5	4	Taste bud	10	224.5	3
Sensory nerve	10	244.5	2	Temperament	20	81.5	6
Sentiment[2]	?	.....	4	Thalamus	14	174.5	7
Set	?	.....	6	Thinking	25	35	7
Sibling	6	305	3	Thymus	5	318	3
Somesthesia	9	263.5	2	Thyroid	16	138.5	8
Sound waves	9	263.5	4	Thyroxin	7	292.5	3
Sour grapes reaction	9	263.5	2	Timbre	10	244.5	6
Span of apprehension	16	138.5	3	Trait	18	112.5	5
Spinal cord	10	244.5	4	Transfer	26	30	7
Standard deviation	19	95.5	3	Traumatic	4	332	1
Stanford-Binet test	20	81.5	5	Trial and error	29	15	7
Startle pattern	5	318	4	Two-point threshold	4	332	2
Statistical	13	192	5	Unconscious	22	60.5	8
Stereoscope	8	280	5	Validity	25	35	3
Stereotype	14	174.5	5	Viscera	14	174.5	2
Stimulus	24	44	8	Visual purple	4	332	4
Strong vocational interest blank	4	332	3	Vitreous humor	3	347.5	1
Subjective	18	112.5	1	Weber's law	15	153.5	5
Sublimation	19	95.5	4	Wish fulfillment	14	174.5	6
Submission	6	305	4	Wishful thinking	20	81.5	3

[1] The range of checks is from 0 to 34; the median term received 14 checks.

[2] Through error the two terms "set" and "sentiment" were listed together as "set sentiment." This probably invalidates the results for these two terms.

not become a hindrance to the process of refinement and clarification of terminology, frequent revision would be imperative.

SUMMARY

A check list of three hundred sixty-six terms found to appear in at least three of eight elementary textbooks in general psychology was sent to one representative psychologist in each of seventy-one uni-

versities and colleges in the United States. Each psychologist was asked to check approximately one hundred terms which in his opinion should be given emphasis in the elementary course. Results obtained from thirty-nine respondents are tabulated in an alphabetized list of the terms. For each term are reported the number of checks received, the rank order for number of checks, and the number of books, of the eight textbooks examined, in which the term is listed in the index. The desirability of a "normative dictionary" of terms important for general psychology is suggested.

BIBLIOGRAPHY

1. Boring, E. G., Langfeld, H. S., and Weld, H. P.: *Introduction to Psychology.* New York: John Wiley and Sons, Inc., 1939, pp. ix and 652.
2. Cole, L. E.: *General Psychology.* New York: McGraw-Hill Company, Inc., 1939, pp. xii and 688.
3. Dashiell, J. F.: *Fundamentals of General Psychology.* Boston: Houghton Mifflin Company, 1937, pp. xxiii and 655.
4. Guilford, J. P.: *General Psychology.* New York: D. Van Nostrand Company, Inc., 1939, pp. xii and 630.
5. Husband, R. W.: *General Psychology.* New York: Farrar and Rinehart, Inc., 1940, pp. xiv and 513.
6. Ruch, F. L.: *Psychology and Life.* New York: Scott, Foresman and Company, 1937, pp. xiv and 679.
7. Tiffin, J., Knight, F. B., and Josey, C. C.: *The Psychology of Normal People.* Boston: D. C. Heath and Company, 1940, pp. xv and 512.
8. Woodworth, R. S.: *Psychology, Fourth Edition.* New York: Henry Holt and Company, 1940, pp. xiii and 639.

GRADES, INTELLIGENCE AND PERSONALITY OF ART SCHOOL FRESHMEN[1]

ALICE I. BRYAN

Assistant Professor, School of Library Service, Columbia University

Approximately one-fourth of the candidates who apply for admission each year to the School of Fine and Applied Arts at Pratt Institute are accepted. The successful candidates are recent high-school graduates with better than average grades who submit superior plates on a series of assigned art problems and give evidence, through character recommendations and personal interviews, that they are socially well-adjusted. During the freshman year, the academic load is heavy and the pace is swift. Students are told repeatedly that they must work diligently, display business-like habits and attitudes and maintain an average grade of 80 if they expect to return the following year. Since ability to earn good grades by consistent production of acceptable work under considerable pressure in a competitive situation is the chief criterion of success in this art school, the present study was concerned with an attempt to measure the extent to which intelligence and personality may contribute to such success.

It was planned also to investigate the relationship between ability to succeed in courses requiring creative art aptitude and ability to perform tasks involving higher mental processes such as are measured by intelligence tests. Differences among students preparing to enter specialized fields of art were also of interest. Such comparisons were possible because, though enrolled in three different departments—Architecture, Art Education and Design—all students take the same foundation art courses during the freshman year. Since the school is co-educational, it was possible to study sex differences in the abilities measured.

Four successive freshmen classes, entering the school from the Fall of 1936 to the Fall of 1939 inclusive, were given the American Council on Education Psychological Examination and the Bernreuter Personality Inventory. The Revised Minnesota Paper Form Board Test was also administered to discover whether ability to perceive spatial relations, as measured by this test, bears any relationship to success in courses in a professional art school. Grades in two major courses, Design and Structural Representation, as well as average grades in

[1] This study was completed with the kind assistance of a grant-in-aid from the Carnegie Corporation.

all art subjects were obtained for all students at the end of their first year of school. Data were obtained for one thousand eight students, of whom four hundred ninety-eight were women and five hundred ten men. Table I shows the distribution of these students, by sex, among the three departments of the school. It should be noted that the students enrolled in Architecture are all men.

TABLE I.—DISTRIBUTION OF 1008 PRATT ART SCHOOL FRESHMEN, BY SEX, ACCORDING TO DEPARTMENT OF SPECIALIZATION

	Architecture[1]	Art Education	Design	Totals
Men	75	61	374	510
Women	..	157	341	498
Total	75	218	715	1008

[1] Records were obtained for only four women enrolled in Architecture. Data on these students were excluded from this study because of the small number of cases.

The tests were administered by the same experimenter to each freshmen group during their first semester. Four different editions of the American Council Examination were used; namely, the 1936, 1937, 1938 and 1939 editions. Scores were available for the entire group on the Completion, Arithmetic and Same-Opposite subtests; since the subtests for Artificial Language and Spatial Analogies were dropped from the 1939 edition, scores on these tests were available for only eight hundred twenty students, four hundred seven men and four hundred thirteen women. Subtests on Number Series in the 1938 and 1939 editions and on Verbal Analogies and Tables in the 1939 edition, although used to compute total scores, were not included as separate variables in this study since less than half the students took the Number Series test and less than a quarter took the other two tests. Because the Architecture and Art Education groups were considerably smaller than the Design group, students enrolling in these two departments in 1939 were given the 1938 edition containing the spatial analogies test which was of special interest for this study. Total scores and scores on all subtests used were converted into percentile ranks according to the Thurstone norms. This procedure was followed in order to make the scores for the successive years comparable. On the assumption that the general college populations were

comparable for the four years, it was possible to equate scores of the same percentile rank.

The Bernreuter Personality Inventory was scored for neurotic tendency (B_1N). All scores were converted into percentile values from the norms for college men and women. Percentile scores for one hundred men and one hundred women, selected at random, were then combined to form a new set of percentile values into which all scores were converted. This was done to facilitate computation of intercorrelations for the combined sex groups with other variables. In analyzing results on the Minnesota Paper Form Board Test, raw scores were used. Grades for the art courses taken by the students, including perspective, drawing, painting, etc., were available in quantitative scores on a scale of one hundred points. The mean of the grades assigned by the teachers of the several subjects constitutes the art grade average used in this study.

TABLE II.—MEANS AND STANDARD DEVIATIONS ON ALL VARIABLES[1] FOR MEN, WOMEN AND TOTAL GROUP

Variables	Means			SD's		
	Men	Women	Total	Men	Women	Total
1. Art Grade Average	81.75	81.17	81.46	5.8	4.9	5.4
2. Art Grade Design	81.41	80.63	81.02	8.5	7.4	8.0
3. Art Grade Structural Representation	82.13	80.41	81.27	7.4	7.6	7.6
4. Minnesota Paper Form Board	44.52	44.52	44.52	9.4	12.4	11.0
5. Bernreuter B_1N	52.53	47.41	49.99	27.5	29.4	28.5
6. American Council Total	43.54	53.74	48.57	26.9	26.6	27.2
7. Completion	49.76	53.95	51.83	29.1	27.9	28.6
8. Arithmetic	41.86	35.44	38.68	27.4	25.3	26.6
9. Same-Opposite	48.36	59.24	53.73	29.2	27.5	28.9
10. Artificial Language	34.89	55.88	45.48	25.2	27.7	28.5
11. Analogies (Spatial)	54.53	59.83	57.22	30.2	27.1	28.8

[1] For Variables 1–9 inclusive, N = 510 Men and 498 Women; for Variables 10 and 11, N = 407 Men and 413 Women.

Means and standard deviations on eleven variables for all men, all women, and the total group are given in Table II. In Table III are shown means and standard deviations for groups of students enrolled in Architecture, Art Education and Design. Table IV shows means

TABLE III.—MEANS AND STANDARD DEVIATIONS FOR STUDENTS ENROLLED IN ARCHITECTURE, ART EDUCATION AND DESIGN

Variables	Architecture[1]		Art Education[2]		Design[3]	
	Means	SD's	Means	SD's	Means	SD's
1. Art Grade Average	78.16	5.7	82.79	4.8	81.40	5.3
2. Art Grade Design	78.57	9.0	82.04	7.0	80.97	8.1
3. Art Grade Structural Representation	78.07	6.5	82.79	6.4	81.16	7.9
4. Minnesota Paper Form Board	44.85	7.8	46.70	17.7	43.82	8.1
5. Bernreuter B_1N	53.67	27.9	47.07	27.9	50.50	28.7
6. American Council Total	40.85	26.3	61.61	24.9	45.41	26.7
7. Completion	44.65	27.4	63.85	25.8	48.92	28.5
8. Arithmetic	54.41	26.9	44.62	26.7	35.23	25.6
9. Same-Opposite	38.43	28.6	62.21	26.9	52.75	28.7
10. Artificial Language	27.65	20.9	62.77	28.2	40.84	26.1
11. Analogies (Spatial)	56.84	32.2	58.04	26.3	56.90	29.3

[1] N = 75 for all Variables.
[2] N = 218 for all Variables.
[3] N = 715 for Variables 1–9 inclusive; for Variables 10 and 11, N = 527.

TABLE IV.—MEANS AND STANDARD DEVIATIONS FOR MEN AND WOMEN ENROLLED IN ART EDUCATION AND DESIGN

Variables	Art Education[1]				Design[2]			
	Means		SD's		Means		SD's	
	Men	Women	Men	Women	Men	Women	Men	Women
1. Art Grade Average	83.66	82.46	5.9	4.3	82.15	80.58	5.5	5.0
2. Art Grade Design	82.74	81.77	7.4	6.8	81.75	80.11	8.5	7.6
3. Art Grade Structural Representation	84.51	82.12	7.2	5.9	82.55	79.63	7.3	8.2
4. Minnesota Paper Form Board	45.80	47.05	15.1	18.5	44.24	43.35	8.4	7.8
5. Bernreuter B_1N	49.64	46.08	27.8	27.9	52.77	48.02	27.3	30.0
6. American Council Total	52.26	65.25	27.2	22.9	42.64	48.44	26.6	26.5
7. Completion	61.43	64.79	27.7	25.0	48.88	48.96	29.2	27.7
8. Arithmetic	47.15	43.64	26.7	26.6	38.48	31.66	26.8	23.7
9. Same-Opposite	53.44	65.61	30.1	24.7	49.51	56.31	28.8	28.2
10. Artificial Language	46.16	69.22	29.9	24.6	34.36	47.69	24.3	26.3
11. Analogies (Spatial)	53.51	59.80	28.9	25.1	54.11	59.86	29.9	28.2

[1] For all Variables, N = 61 Men and 157 Women.
[2] For Variables 1–9 inclusive, N = 374 Men and 341 Women; for Variables 10 and 11, N = 271 Men and 256 Women.

and standard deviations for men and women separately in the Art Education and Design groups. The Architecture group was composed of men students only, so no sex comparisons were possible. The sigmas of the means in Table II range from 0.2 to 1.5, in Table III from 0.2 to 3.7, and in Table IV from 0.3 to 3.9. In Table V are given differences between means for students enrolled in Architecture, Art Education and Design with their D/SD(diff.) values. Table VI shows sex differences and D/SD(diff.) values for the group as a whole and for the students enrolled in the Art Education and Design subgroups.

TABLE V.—DIFFERENCES BETWEEN MEANS OF GROUPS ENROLLED IN ARCHITECTURE, ART EDUCATION AND DESIGN[1]

	Architecture and art education		Architecture and design		Art education and design	
	D	D/SD (diff.)	D	D/SD (diff.)	D	D/SD (diff.)
1. Art Grade Average	− 4.63	6.09	− 3.24	4.44	1.39	3.86
2. Art Grade Design	− 3.47	3.10	− 2.40	2.31	1.07	1.84
3. Art Grade Structural Representation	− 4.72	5.83	− 3.09	4.07	1.63	3.26
4. Minnesota Paper Form Board	− 1.85	1.23	1.03	1.08	2.88	2.32
5. Bernreuter B_1N	6.60	1.77	3.17	0.94	− 3.43	1.56
6. American Council Total	−20.76	6.02	− 4.56	1.44	16.20	8.22
7. Completion	−19.20	5.23	− 4.27	1.26	14.93	7.08
8. Arithmetic	9.79	2.73	19.18	5.88	9.39	4.56
9. Same-Opposite	−23.78	6.32	−14.32	4.11	9.46	4.48
10. Artificial Language	−35.12	11.48	−13.19	5.00	21.93	9.97
11. Analogies (Spatial)	− 1.20	0.29	− 0.06	0.02	1.14	0.51

[1] Positive entries in the D columns indicate differences in means in favor of the first group of each pair listed; negative entries in favor of the second group. On Variable 5, a higher mean score denotes greater neurotic tendency.

With respect to intelligence, as measured by the American Council test, it seems clear from Table II that this group of art school freshmen taken as a whole is fairly typical of the general college population, the mean for the entire group falling at the 49th percentile. That a selective factor is operating, however, with regard to the intelligence level of men and women students is apparent from the fact that the mean score for the men is at the 44th percentile, while for the women

the mean is at the 54th percentile. If a D/SD(diff.) of 2.50 is accepted as an indication of reliability, this difference of ten percentiles is highly reliable since the D/SD(diff.) is 6.00, as shown in Table VI. This superiority of the women students holds for the Art Education and Design subgroups, as well, although the difference is more marked among the Art Education students. The mean for the women in the Art Education group is 12.99 percentiles higher than that of the men, while the mean for the women in the Design group is 5.80 percentiles higher than the men's; the D/SD(diffs.) are 3.30 and 2.93, respectively.

TABLE VI.—SEX DIFFERENCES FOR TOTAL GROUP AND FOR ART EDUCATION AND DESIGN GROUPS[1]

	Total group		Art Education		Design	
	D	D/SD (diff.)	D	D/SD (diff.)	D	D/SD (diff.)
1. Art Grade Average	.58	1.61	1.20	1.41	1.57	3.74
2. Art Grade Design	.78	1.56	0.97	0.94	1.64	2.88
3. Art Grade Structural Representation	1.72	4.09	2.39	2.32	2.92	5.12
4. Minnesota Paper Form Board	.00	.....	− 1.25	0.52	0.89	1.56
5. Bernreuter B_1N	5.12	2.89	3.56	0.84	4.75	2.23
6. American Council Total	−10.20	6.00	−12.99	3.30	− 5.80	2.93
7. Completion	− 4.19	2.36	− 3.36	0.83	− 0.08	0.04
8. Arithmetic	6.42	3.93	3.51	0.88	6.82	3.57
9. Same-Opposite	−10.88	6.14	−12.71	2.78	− 6.80	3.21
10. Artificial Language	−20.99	10.98	−23.06	5.38	−13.33	6.09
11. Analogies (Spatial)	− 5.30	2.66	− 6.29	1.49	− 5.75	2.25

[1] Positive entries in the D columns indicate differences in means in favor of the men; negative entries in favor of the women. On Variable 5, a higher mean score denotes greater neurotic tendency.

As a group, the Architecture students, all of whom are men, are considerably inferior to the Art Education group, as shown in Table V. The difference in means is 20.76 percentiles with a critical ratio of 6.02. Comparing the male Architecture students with the male Art Education group, there is a difference in favor of the latter of 11.41 percentiles with a critical ratio of 2.48. The women in Art Education show a mean score superiority of 24.40 percentiles, with a critical ratio of 6.97, over the men in Architecture. Compared with the Design students, the Architecture group is lower in mean

score by 4.56 percentiles, but this difference is not reliable, the critical ratio being only 1.44. The difference between the men in Design and in Architecture is only 1.79 in favor of the Design group, with a D/SD(diff.) of but .54. It is apparent that the superiority of the group of women as a whole over the men as a whole is mainly due to the fact that the women in Art Education are a decidedly superior group, while the men in Architecture are below the mean for college freshmen in general. An explanation for this situation may be found in the tendency of male architects who can afford to do so to acquire a liberal arts background before beginning their professional training. This tendency is probably not so prevalent in women preparing to become art teachers.

The American Council subtests of verbal ability indicate likewise that the Pratt freshmen are a fairly normal group compared with liberal arts freshmen. On the Completion test, the mean for the entire group is at the 52nd percentile; on the Same-Opposite test, at the 54th percentile; and on the Artificial Language test, at the 46th percentile. The means for the women are higher on all three of these tests. On Completion, this difference is 4.19 percentiles; on Same-Opposite, 10.88 percentiles; and on Artificial Language, 20.99 percentiles. The D/SD(diffs.) are 2.36, 6.14, and 10.98, respectively, indicating a highly reliable superiority for the women on the Same-Opposite and Artificial Language tests. Again, the women in Art Education show a greater superiority over the men on these verbal tests than do the women in Design. The differences between men and women in these subgroups, as shown in Table VI, are significant for the Same-Opposite and Artificial Language tests, but not for Completion.

On the Arithmetic test, the mean for the whole Pratt group (Table II) is at the 39th percentile, considerably below the mean for liberal arts students. This is the only subtest of the American Council Examination on which the women as a whole scored lower than the men. The mean for the men is 6.42 percentiles higher, with a D/SD(diff.) of 3.93. The men students in Design show a reliable superiority of 6.82 percentiles over the women, while the men in Art Education are superior by only 3.51 percentiles, a difference which is not reliable. This is also the only test on which the Architecture students are superior to the other two groups. They show a reliable difference in mean score on Arithmetic of 9.79 percentiles over the Art Education students and a difference of 19.18 over the Design group. The Art

Education students are reliably higher than the Design students by 9.39 percentiles on the Arithmetic test.

Both men and women of the entire group, as well as all three subgroups, (Tables II and III), show a higher than average score on the Spatial Analogies subtest. The mean for the group as a whole is at the 57th percentile; for all men combined, at the 55th; for all women combined, at the 60th percentile. The mean for the Architecture group falls at the 57th percentile; for the Art Education group at the 58th; and for the Design group at the 57th percentile. The women as a whole are reliably superior on this test to the men by 5.30 percentiles; the women in Art Education are superior to the men, but not reliably so, by 6.29 percentiles; and the women in Design are superior to the men by 5.75 percentiles, a difference which also is not reliable. Differences between the Architecture, Art Education and Design groups are not reliable and are less than two percentiles in every case. This consistent superiority of all groups of Pratt Art freshmen, both men and women, on the Spatial Analogies test suggests that this test alone of the American Council battery may have some significance in differentiating college freshmen with art ability from the general population of liberal arts college students.

An equally consistent superiority of Pratt Art School groups in ability to perceive spatial relations is indicated by their mean scores on the Minnesota Paper Form Board Test (Tables II and III). The mean for the group as a whole is 44.52, with means for all men combined and all women combined falling also at 44.52. The median scores reported by the authors of this test for a group of two hundred forty-seven freshmen in the College of Arts and Pure Science at New York University and for a group of one thousand two hundred eighty-eight high-school seniors applying for admission to this college and to the College of Engineering at New York University are 39 and 40 respectively. The mean of the Pratt students falls at the 75th percentile of these two groups. Again, on this test, the differences between the Architecture, Art Education and Design groups at Pratt are less than 3 percentiles and are not reliable. Means for the men and women in Art Education and Design likewise differ by less than 1.5 percentiles.

On the Bernreuter Personality Inventory, scored for neurotic tendency, the mean for the total group at Pratt (Table II) falls at the 50th percentile of college students in general. The mean of 52.53 for the men is 5.12 percentiles above that of 47.41 for the women, with a

D/SD(diff.) of 2.89, indicating a slightly higher neurotic tendency on the part of the men. This difference, while reliable, is so small that it is probably of no educational significance, particularly since the means for both men and women are so close to normality. The Architecture group has a mean which is 6.60 percentiles higher than the Art Education group and 3.17 percentiles higher than the Design group (Table III), but neither of these differences is reliable. Differences between men and women in the Art Education and Design groups are less than 5 percentiles and are not statistically reliable. It seems evident that, insofar as the Bernreuter Inventory is a valid measure of neurotic tendency, these one thousand eight art school students are no more and no less neurotic than college freshmen in general.

The mean for average grades in art courses for the total Pratt group, as shown in Table II, is 81.46 per cent. Means for art grades in the Design and Structural Representation courses are likewise about 81 per cent. Sex differences in grades for the whole group are so small as to be negligible. Students enrolled in Architecture show small and for the most part reliable differences in the direction of inferiority to the Art Education and Design groups. This inferiority is slightly greater and more reliable for the course in Structural Representation than for the course in Design. The Art Education group is slightly but unreliably superior in grades to the Design group. Differences between men and women students in Art Education and Design favor the women, but are very small, the largest difference being 2.92 with a *D*/SD(diff.) of 5.12 between men and women in the Design group.

Summarizing the results shown in Tables II to VI inclusive, it may be said that the Pratt Art School freshmen as a total group are comparable in general intelligence, as measured by the American Council test, to the larger population of college freshmen, although they display a weakness in arithmetical reasoning and a compensating strength in perceiving spatial analogies. Their superiority over liberal arts freshmen in ability to perceive spatial relations is further demonstrated by reliably higher scores on the Minnesota Paper Form Board Test. With respect to neurotic tendency, popularly attributed to individuals with artistic inclinations, the Pratt students appear to be just as stable and well-adjusted as the average liberal arts college freshman, as far as this trait can be measured by the Bernreuter Personality Inventory. In comparing groups of students enrolled in

the three departments of the school, it is evident that those preparing to become art teachers are reliably superior to the other groups and to the average college freshman in practically every trait measured. This finding is at variance with results of studies, such as that conducted by Learned in Pennsylvania, which stress the inferiority of teacher-training students in general as compared with liberal arts groups.

TABLE VII.—INTERCORRELATIONS OF ALL VARIABLES FOR TOTAL GROUP, MEN AND WOMEN COMBINED

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average		.68	.62	.19	−.02	.16	.16	.10	.15	.05	.11
2. Design			.40	.19	−.02	.10	.13	.11	.07	−.02	.06
3. Structural Representation				.11	−.01	.11	.11	.06	.10	.08	.06
4. Minnesota					−.08	.29	.19	.28	.14	.21	.27
5. Bernreuter						−.08	−.03	−.02	−.02	−.03	−.08
6. American Council Total							.75	.57	.79	.72	.63
7. Completion								.40	.63	.46	.32
8. Arithmetic									.30	.30	.37
9. Same-Opposite										.44	.31
10. Artificial Language											.36
11. Analogies											

Intercorrelations among the eleven variables for the total group, for all men combined, and for all women combined are given in Tables VII and VIII.[1] In Tables IX and X are shown the intercorrelations obtained for the groups of students enrolled in Architecture, Art Education and Design, treated separately. Tables XI and XII give intercorrelations for the men and women students enrolled in Art Education and Design. Probable errors were computed for the high and low coefficients of correlation for all groups (Tables VII through XII). All were found to be negligible for correlations of any significance, the range being from .00 to .06.

As might be expected, correlations of the American Council total score with the subtests in this battery, shown in Table VII, are reliably high, ranging from .57 for Arithmetic to .79 for the Same-Opposite test. Within the subgroups, the lowest correlation found for the total

[1] For the equivalence of correlations between raw scores and percentiles, see Bennett, George K.: "The correlation between scores expressed as percentiles." *Jour. Educ. Psych.*, Sept. 1939.

score was .52 with Artificial Language in the Architecture group; while the highest was .86 with Artificial Language in the group of all women combined. Among the subtests of the American Council

TABLE VIII.—INTERCORRELATIONS OF ALL VARIABLES FOR ALL MEN COMBINED AND ALL WOMEN COMBINED
(Men above Diagonal, Women below)

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average	.....	.65	.58	.15	.00	.13	.11	.07	.12	.07	.15
2. Design	.68	.....	.37	.19	.00	.13	.12	.12	.08	.08	.20
3. Structural Representation	.63	.40	.....	.08	.01	.11	.09	.03	.08	−.02	−.02
4. Minnesota	.21	.19	.12	.....	−.09	.28	.18	.27	.11	.30	.45
5. Bernreuter	−.06	−.07	−.06	−.08	.....	−.02	.01	−.03	.02	.01	−.06
6. American Council Total	.22	.10	.15	.31	−.04	.....	.75	.62	.77	.75	.72
7. Completion	.23	.16	.15	.20	−.06	.76	...	.41	.63	.58	.46
8. Arithmetic	.13	.09	.07	.31	−.04	.60	.43	.....	.41	.53	.68
9. Same-Opposite	.21	.13	.17	.17	−.03	.79	.63	.36	...	.43	.30
10. Artificial Language	.23	.16	.17	.39	−.05	.86	.61	.65	.46	.....	.34
11. Analogies	.24	.22	.04	.39	−.10	.74	.48	.58	.37	.38	

TABLE IX.—INTERCORRELATIONS OF ALL VARIABLES FOR STUDENTS ENROLLED IN ARCHITECTURE

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average	...	.74	.42	.33	−.03	.37	.27	.36	.30	.03	.30
2. Design	...	...	.25	.25	−.15	.22	.20	.28	.16	.06	.17
3. Structural Representation	...	...	...	.17	.00	.06	.01	.12	−.02	−.09	.16
4. Minnesota	...	...	...	...	−.13	.36	.14	.25	.15	.20	.43
5. Bernreuter	...	...	...	...	.....	−.17	.00	−.24	−.04	−.13	−.29
6. American Council Total	...	...	...	...	.....	.....	.71	.69	.71	.52	.72
7. Completion	...	...	...	...	.....	.....	...	.48	.52	.29	.33
8. Arithmetic	...	...	...	...	.....	.....	...	.....	.33	.40	.53
9. Same-Opposite	...	...	...	...	.....	.....	...	.....	.....	.25	.30
10. Artificial Language	...	...	...	...	.....	.....	...	.....	.....	.....	.26
11. Analogies	...										

battery, the lowest correlation was .16 between the Same-Opposite and Spatial Analogies tests for the men enrolled in Art Education; the highest coefficient was .72 between Arithmetic and Spatial Analogies for the men enrolled in Design.

Correlations between the American Council total score and the Minnesota Paper Form Board test ranged from .05 for the men in Art Education to .36 for the men in Architecture. Among the subtests of the American Council scale, the lowest correlation with the Minnesota test was an inverse relationship of −.06 with the Same-Opposite test for the men in Art Education; while the highest was .56 with Spatial Analogies for the women in the Design group. Fairly high correlations between the Minnesota test and the Spatial Analogies test included coefficients of .55 for the total Design group, .55 for the men in the Design group, .45 for all men combined, .43 for the Architecture group, and .39 for all women combined.

TABLE X.—INTERCORRELATIONS OF ALL VARIABLES FOR STUDENTS ENROLLED IN ART EDUCATION AND DESIGN
(Art Education Group above Diagonal, Design below)

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average	...	.59	.77	.17	−.06	.05	.13	.04	.12	−.09	.08
2. Design	.69	.....	.40	.21	.04	.02	.11	.06	.05	−.02	.01
3. Structural Representation	.59	.40	.....	.15	−.02	.06	.06	.06	.04	−.08	.11
4. Minnesota	.20	.20	.09	.....	−.11	.21	.16	.23	.06	.20	.22
5. Bernreuter	.01	−.02	.00	−.06	.....	−.11	−.13	−.06	−.09	−.04	−.16
6. American Council Total	.12	.09	.09	.34	.00	.....	.74	.54	.77	.72	.54
7. Completion	.11	.11	.10	.21	.01	.74	.....	.34	.62	.38	.33
8. Arithmetic	.12	.12	.07	.34	.01	.59	.42	.....	.28	.33	.35
9. Same-Opposite	.09	.04	.08	.19	.02	.80	.63	.34	.....	.34	.21
10. Artificial Language	.09	.09	−.02	.39	−.01	.78	.58	.56	.49	.....	.36
11. Analogies	.20	.27	−.07	.55	−.07	.76	.49	.65	.39	.42	

Relationships between average art grades and grades in the courses in Design and Structural Representation (Table VII) were comparable in magnitude with correlations between the American Council total score and its subtests. Correlations between average art grades and grades in the Design course ranged from .59 for the Art Education group to .74 for the Architecture group. Correlations between average art grades and grades in Structural Representation fell between .42 for the Architecture group and .81 for the men in Art Education.

Between average art grades and American Council total scores, coefficients of correlation were obtained ranging from .02 for the men

in Art Education to .37 for the Architecture group, the second highest correlation amounting to .22 for all women combined. Among the subtests of the American Council, correlations were obtained with average art grades which ranged from −.10 with Artificial Language for the men in the Art Education group to .36 with Arithmetic for the Architecture students. Between the Minnesota test and average art grades, correlations were obtained ranging from .10 for the Art Education group to .33 for Architecture.

TABLE XI.—INTERCORRELATIONS OF ALL VARIABLES FOR MEN AND WOMEN ENROLLED IN ART EDUCATION
(Men above Diagonal, Women below)

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average	.....	.60	.81	.10	−.09	.02	.00	−.03	.02	−.10	.08
2. Design	.59	...	.41	.22	−.02	.02	−.01	.00	.08	−.08	−.04
3. Structural Representation	.74	.39	.....	.12	−.01	.19	.07	.13	.08	.03	.29
4. Minnesota	.21	.21	.18	.....	−.05	.05	.06	.08	−.06	.00	.21
5. Bernreuter	−.06	.06	−.04	−.12	.....	−.03	−.08	.14	−.04	.03	.00
6. American Council Total	.13	.05	.05	.28	−.13	.....	.73	.53	.82	.70	.49
7. Completion	.21	.17	.08	.20	−.14	.77	.....	.22	.61	.45	.25
8. Arithmetic	.07	.08	.01	.29	−.15	.59	.40	.....	.35	.30	.38
9. Same-Opposite	.23	.05	.08	.10	−.10	.73	.62	.28	.....	.35	.16
10. Artificial Language	−.02	.04	−.06	.29	−.04	.71	.36	.41	.26	.....	.34
11. Analogies	.10	.04	.05	.23	−.22	.56	.37	.36	.20	.35	

Correlations from −.08 to .30 were obtained between grades in the Design course and the American Council total score and subtests, the highest correlation being with Spatial Analogies for the men in the Design group. A range of correlations from .19 to .25 was found between Design grades and the Minnesota test, the highest correlation occurring in the Architecture group. Between grades in Structural Representation and the American Council battery, correlations ranged from −.09 with Artificial Language for the Architecture students to .29 with Spatial Analogies for the men in Art Education. The lowest correlation obtained between Structural Representation grades and the Minnesota test was .08, the highest .18 for the women in Art Education.

Intercorrelations of the Bernreuter test with all other variables for all groups as shown in Tables VII to XII, ranged from −.29 with

Spatial Analogies for the Architecture students to .14 with Arithmetic for the men in Art Education. Between average art grades and Bernreuter, correlations ranged between −.09 for the men in Art Education to .05 for the men in Design. For grades in Design and Bernreuter, a range of coefficients from −.15 for Architecture to .06 for women in Art Education was found. Between grades in Structural Representation and Bernreuter, correlations ranged from −.07 for women in the Design group to .03 for men in the Design group.

TABLE XII.—INTERCORRELATIONS OF ALL VARIABLES FOR MEN AND WOMEN ENROLLED IN DESIGN
(Men above Diagonal, Women below)

	1	2	3	4	5	6	7	8	9	10	11
1. Art Average	.....	.67	.58	.16	.05	.09	.06	.10	.06	.09	.21
2. Design	.70	.....	.39	.20	.05	.12	.11	.14	.05	.11	.30
3. Structural Representation	.58	.39	.....	.08	.03	.10	.07	.06	.06	−.07	−.07
4. Minnesota	.23	.19	.08	.....	−.09	.34	.23	.34	.16	.44	.55
5. Bernreuter	−.06	−.11	−.07	−.04	.....	.01	.03	−.01	.05	.06	−.08
6. American Council Total	.19	.08	.13	.35	.00	...	.76	.64	.78	.79	.76
7. Completion	.18	.13	.13	.20	−.02	74	...	.44	.65	.65	.54
8. Arithmetic	.11	.06	.04	.33	.01	.58	.39	.....	.35	.61	.72
9. Same-Opposite	.17	.05	.17	.24	.01	.81	.62	.36	...	.49	.37
10. Artificial Language	.19	.13	.13	.39	−.03	.77	.55	.65	.45	.....	.38
11. Analogies	.24	.27	−.02	.56	−.03	.74	.44	.63	.38	.45	

Results obtained from computation of intercorrelations among the eleven variables of this study may be summarized as follows: Reliably high correlations were found between the total score and the subtests of the American Council battery and between the average art grades and the grades in the Design and Structural Representation courses. Positive but, for the most part, low correlations were obtained between the American Council total score and the Minnesota Paper Form Board Test, the highest correlation being .36 for the Architecture students. The highest correlations between the Minnesota test and the American Council subtests were found for the test on Spatial Relations, a coefficient of .55 for the Design group representing the greatest degree of relationship.

Art grades were related to scores on the American Council test in a positive direction, but correlations were so low as to have little

value for predictive purposes. The highest correlation obtained was .37 between average art grades and total score on American Council; this was found for Architecture students whose mean on this test was reliably lower than that of any other group. The Minnesota Paper Form Board test appeared to have no greater predictive value for success in art courses than did the American Council test.

Correlations between grades in art courses and stability of personality as measured by the Bernreuter Inventory were also so low as to indicate little or no relationship.

Findings of this study suggest that success of freshmen students in Pratt Institute Art School depends upon creative abilities other than those measured by paper-and-pencil tests of intelligence and personality.

SYMBOLICAL PRACTICE

J. B. STROUD AND MAX FREEBURNE

State University of Iowa

INTRODUCTION

In the past symbolical practice has been studied in connection with learning trials, as when a certain number of recall trials without the copy are substituted for reading trials or other modes of presentation, and in review trials, as when reviewing is done by direct recall or other means not involving representation of the copy.* The problem at hand pertains to the question of the relationship between the effectiveness of symbolical practice, in comparison with the non-symbolical, and the stage of learning and forgetting at which it is introduced. Skaggs, Grossman, Krueger and Krueger[5] found that in a series of trials recall was relatively ineffective when introduced near the beginning, that is, within the first three or four trials. Lyon[2] recommended that review, by which he seems to have meant direct recall and recall aided by inspection of lecture notes, should be engaged in relatively soon after learning—the same day, ideally, and that of several such review exercises the initial ones should be rather closely spaced.

Recent work has shown that Lyon's recommendation is well founded in so far as objective test exercises function as review media.[6,7,8] Incidentally it has been found that the multiple choice form of item has the same general effectiveness in this respect as does the completion form.[9] The former is commonly regarded as a recognition item (although undoubtedly some recall is often present) and the latter, as a recall item. It is suggested, therefore, that an act of recognition is conducive to learning and retention just as is an act of recall. Perhaps Raffel's work comes closest of any to being a direct test and verification of Lyon's pronouncement.[4] One of her groups was given a recall exercise daily for seven days. This group achieved on the seventh day a score ninety per cent of the original score. A group not having intervening recall trials achieved a mean score sixty-five per cent of the original. A group that engaged in recall trials on the second, fifth, and sixth days achieved a final score eighty per cent of the original.

* Thus the term symbolical practice refers to what is known as recitation or recall in learning trials and recall in review trials.

From the foregoing it has seemed that the relative effectiveness of symbolical trials both in primary learning and review should be a function of the command which the subject has of the material. In the initial learning trials when the subject has gained but little mastery of the material, non-symbolical trials should be superior to symbolical trials. Likewise, as forgetting progresses to the point that much of the material is beyond symbolical representation, review by reimpression should prove to be more efficacious than review by recall or other such means. If this condition prevails, its ascertainment should have some practical value.

The purpose of the present article is to describe an investigation carried out with the view to testing the foregoing assumptions. The investigation has as a secondary purpose the making of a further test of a point previously raised;[2,6,7] namely, that review should follow learning rather closely if the review is symbolical in character, but that the effectiveness of non-symbolical review is largely independent of the length of time elapsing between the learning and review exercises—within the limits studied. In the investigations in which this question has been studied directly, objective tests served as symbolical review media. In the study reported here recall is the symbolical medium. There are also other important differences; to wit, in the former studies logical or substance learning constituted the activity of the subjects and, as is customary with such materials, learning was not carried to a criterion; all subjects were given the same learning time; that is, practice time was constant and amount of learning varied; scores on objective tests indicated the achievement of the subjects. In the present study learning was carried to a criterion; that is, amount was constant and practice time varied. The learning activity was memorization rather than substance learning.

PROCEDURE

The subjects in this experiment were thirty-two college students, fourteen graduate and eighteen undergraduate. Each subject learned four lists of logically unrelated pairs of adjectives, fourteen pairs per list, presented by means of a Hull drum, the exposure period being two seconds. The pairs in each list were presented in three different serial orders. The exact procedure to be followed in the presentation of the material was explained to the subjects. They were also given a practice list which was learned to a criterion of fifty per cent correct, presentation and recall trials alternating. Four conditions of learn-

ing and four conditions of review were utilized in the experiment, each subject participating in all conditions. A preliminary experiment demonstrated the inadvisability of introducing recall trials after a predetermined number of presentation trials, owing to the fact that such a procedure introduced recall at an early stage for some subjects and at an intermediate or late stage for others, depending upon their respective rates of learning. It was determined to carry each subject to a criterion of one correct repetition and to change the proportion of recall and presentation trials after a given number of items had been learned.

It is, of course, impossible for a subject to master a list by recall trials alone; and while mastery is possible by presentation trials alone, it is impossible to know when mastery has been attained, when the method of direct recall is used as the experimental procedure. Therefore, tests of the comparative effectiveness of recall and presentation trials are made by varying the proportion of each. In the investigation at hand the proportion was two to one, as is seen in the following four conditions. In Condition I the trials were allowed in the proportion of two presentations and one recall until one-half of the pairs in a list had been learned (one-half or more), and thereafter until all pairs were learned, in the proportion of one presentation and two recall trials. Condition II is the reverse of Condition I in that in the acquisition of the first one-half of the pairs the trials were in the proportion of one presentation and two recalls, and in the last half, two presentations and one recall. In Condition III, trials were imposed in the proportion of two presentations and one recall throughout, and in Condition IV, one presentation and two recalls.

In this experiment presentation was made by simultaneous exposure of each pair, the subjects reading the pairs silently. Recall trials were made by exposing the first adjective only of each pair, the subject naming audibly, if he could, its associate. Inter-condition practice effects and probable differences in difficulty of the lists were controlled by counterbalancing. That is, the order of the conditions was so varied that each appeared first, second, third, and fourth an equal number of times and each list appeared in each condition an equal number of times.

For the second objective of this experiment, that of determining the influence of length of interval between learning and review upon the relative effectiveness of recall and presentation trials as a review media, four conditions were utilized. In Condition *A* five presentation

trials (without any recall trials) were imposed one day following the learning exercise. In Condition *B* five recall trials (without any presentation trials) were likewise imposed one day following the learning exercise. Condition *C* allowed five presentation trials fourteen days after learning; and *D*, five recall trials fourteen days after learning. Retention tests, which consisted of three consecutive recall trials without any intervening presentations, were made two days after each review condition. This arrangement makes possible comparison of two kinds of review at each of two temporal positions. Each subject completed one learning and one review condition before being engaged in another.

RESULTS

The reader will recall that in Condition I trials were distributed in the proportion of two presentations and one recall until one-half of the items were learned, and thereafter in the proportion of one presentation and two recalls until the criterion of one perfect repetition was achieved. Condition II was the reverse of Condition I. In Condition III the trials were given in the proportion of two presentations and one recall throughout; and in Condition IV, one presentation and two recalls. The results are shown as follows:

	Condition				Critical ratio
	I	II	III	IV	
Trials to reach one-half criterion	7.5	9.8	9.3	10.1	I II = 2.5; III IV = .7
Additional trials to reach criterion	18.5	16.1	18.8	18.4	I II = .9; III IV = .2

In Conditions I and II a numerical advantage goes to the arrangement providing two presentations to one recall both in the learning of the initial one-half of the items and in the completion of the list. The differences are not statistically significant. Moreover, there was an absence of a numerical difference between Conditions III and IV, either in the learning of the first half of the items or in the learning of the last half.

On the first three trials in Condition III, two presentations and one recall, the average score was reliably higher than that in Condition IV,

one presentation and two recalls (CR = 3.3). At the end of the sixth trial the advantage favored Condition III, also, but not reliably so. Thus only in the very earliest stage of learning was a significant difference obtained in favor of either arrangement, and that, the arrangement providing the greater proportion of presentation trials. At no other stage in the course of learning did one arrangement have any significant advantage over the other.

These results are not to be construed as a refutation of Gates,[1] because the conditions of the experiment are different in certain important respects from those of his investigation; but they do bring into question the generality of his results. Educational psychologists, the writer included, have not described Gates's experiment very well. They have given the impression that his subjects utilized more time in recall than they did. For example, it is customarily stated that under one condition one-fifth of the learning time was devoted to recall; under another, two-fifths; and so on up to four-fifths. Actually this is not the case. Recall was permitted after one-fifth, two-fifths, etc. of the time had elapsed. When recall was introduced, the subjects were instructed to recite sub-vocally as much as they could and to refer to the copy as often as required. We do not know how much of the learning time was actually devoted to recall. He demonstrated the value of recall; but it is possible that recall in the amount commonly claimed on the basis of his results is not profitable. It is also very probable that the effectiveness of symbolical practice varies with a number of conditions as yet untested experimentally.

In the present study the results with respect to the review conditions, subject to the limitations of the experiment, are decisive. In Condition *A* the subjects were given five recall trials as review one day after the primary learning exercise. Two days afterward they achieved a mean retention score of 9.4. In Condition *B* they were given five presentation trials as review likewise one day after primary learning, and two days afterward achieved a mean retention score of 10.4. In Condition *C*, in which five recall trials were imposed fourteen days after primary learning, the mean retention score was 3.1; and in Condition *D*, in which five presentation trials were imposed fourteen days after primary learning, the average score was 8.1. In Conditions *C* and *D*, as in *A* and *B*, the retention test was made two days following the review trials. The difference between the means for *C* and *D* is 6.2 times the standard error of the difference.

SUMMARY

The results of the second aspect of the problem clearly substantiate Lyon in his claim that review by recall should for the best results come soon after learning. In previous experiments in logical or substance learning, experiments in which mastery was not obtained and in which retention was measured by a less exacting method than that employed here, no advantage has been found in introducing non-symbolical reviews (rereading) at an early period as opposed to a late one. In this investigation a difference of 2.3 score points is found in favor of the induction of such reviews on the first day rather than on the fourteenth day, the CR being 2.8. Thus there is a suggestion in these results that, for the retention of memorized material, there may also be some advantage in introducing non-symbolical review at an early date. It is also observed that review by recall trials under the conditions of this experiment is not found to be superior to review by presentation trials when introduced the day following the learning exercise. In this respect the results differ from those of experiments that have dealt with substance learning.

Under the conditions of this experiment no advantage is found in favor of recall trials at any point. The assumption that the comparative effectiveness of recall and presentation trials is a function of the proportion of items that can be recalled—of the degree of mastery attained or of the amount forgotten—is supported in part. Presentation trials are definitely superior to recall trials in the very earliest stages of learning and in review after a lapse of two weeks. Contrary to expectation a preponderance of recall trials did not prove to be superior to a preponderance of presentation trials at any stage of learning. With different proportions the expected advantage in recall trials at the advanced stage of learning might have been realized.

BIBLIOGRAPHY

(1) Gates, A. I.: "Recitation as a factor in memorizing." *Archives of psychology*, Vol. VI, No. 40, 1917.

(2) Lyon, D. O.: "The relation of length of material to time taken for learning." *Journal of educational psychology*, Vol. V, 1914, pp. 1–9, 85–91, 155–163.

(3) Peterson, H. A., Ellis, M., Toohill, N., and Kloess, P., "Some measurements of the effects of reviews." *Journal of educational psychology*, Vol. XXVI, 1935, pp. 65–72.

(4) Raffel, G., "The Effect of recall on forgetting." *Journal of experimental psychology*, Vol. XVII, 1934, pp. 828–838.

(5) Skaggs, E. B., Grossman, S., Kreuger, L., and Kreuger, W. C. F.: "Further studies of the reading-recitation process in learning." *Archives of psychology*, Vol. XVIII, No. 114, 1930.

(6) Sones, A. M., and Stroud, J. B.: "Review, with special reference to temporal position." *Journal of educational psychology*, Vol. XXXI, 1940, pp. 665–676.

(7) Stroud, J. B., and Johnson, E.: "The temporal position of reviews." *Journal of educational research* (in press).

(8) Spitzer, H. F.: "Studies in retention." *Journal of educational psychology*, Vol. XXX, 1939, pp. 641–656.

(9) Wood, W. F.: "The comparative effects of recognition and recall on retention." Master's thesis, State University of Iowa, 1941.

AN ANALYSIS OF THE RANGE OF TESTING AND SCATTERING AMONG RETARDED CHILDREN ON FORM M OF THE REVISED STANFORD-BINET SCALE*

ARTHUR BERGER AND MORRIS SPEEVACK

Psychologists in the Bureau for Children with Retarded Mental Development, Board of Education, New York City

In a previous study[1] the writers made an analysis of the range of testing and scattering on Form L of the Revised Stanford-Binet Scale. The necessity of extending the scale beyond the first-year level in which the subject failed all the test items was indicated, since forty-two per cent of the cases studied increased on the average of 3.2 months of mental age when the test was thus extended.

Since the two scales, L and M, "differ almost completely in content, but are mutually equivalent with respect to difficulty, range, reliability and validity,"[2] the question arises as to whether a change in the content of the test items would affect the range of testing and scattering. It is the purpose of this investigation to study the scatter on Form M and to compare the findings with those of the previous study on Form L.

The procedure of the present study parallels that of the previous one in every respect. A random sampling of two hundred four cases of children in elementary school who were referred for examination because of suspected mental retardation was taken. The chronological age range of the group was from seven years, zero months to fifteen years, eleven months. Table I indicates the number of subjects according to age.

In the administration of Form M, the test was extended to two year levels in which the subject failed all the items. In many cases, the two years of failures were consecutive. Where successes followed a year level in which all items were failed, the test was extended until the "second zero point" was reached.

A comparison was made between the mental ages obtained under the standard procedure and the mental ages obtained when the test

* This supplements a previous analysis of Form L.

[1] Berger, A. and Speevack, M.: "An Analysis of the Range of Testing and Scattering Among Retarded Children on Form L of the Revised Stanford-Binet Scale." *Journal of Educational Psychology*, Vol. XXXI, 1940, pp. 39–44.

[2] Terman, L. and Merril, M.: *Measuring Intelligence.* Houghton-Mifflin Co., Boston, 1937, p. 3.

was extended. Table II shows, according to age, the range of increase in mental age when the test was continued beyond the first year of failure on all test items.

TABLE I.—STANDARD AND CORRECTED IQ RANGE OF TWO HUNDRED FOUR PUPILS REFERRED FOR EXAMINATION ACCORDING TO AGE

Age	IQ range (corrected)	Mean IQ (corrected)	IQ range (standard)	Mean IQ (standard)	Number of cases
7–0 to 7–11	68–80	74.88	68–79	74.25	8
8–0 to 8–11	66–96	75.17	63–96	74.50	30
9–0 to 9–11	61–87	72.57	59–87	71.52	44
10–0 to 10–11	58–98	72.44	58–97	71.29	34
11–0 to 11–11	53–85	69.92	50–83	69.17	36
12–0 to 12–11	54–88	69.22	54–88	68.61	23
13–0 to 13–11	54–70	61.44	54–70	61.00	9
14–0 to 14–11	52–83	67.00	51–83	66.72	18
15–0 to 15–11	53–72	62.50	52.65	58.50	2
Total.....	.....	.....	.....	.....	204

TABLE II.—DIFFERENCES IN MENTAL AGE OBTAINED WHEN TEST IS EXTENDED BEYOND THE FIRST ZERO POINT

Age	All cases		Cases showing change in MA			
	Number	Range of increase in months of MA	Average increase in months of MA	Number	Range of increase in months of MA	Average increase in months of MA
7–0 to 7–11	8	0–2	0.5	2	2	2.0
8–0 to 8–11	30	0–4	0.7	8	2–4	2.8
9–0 to 9–11	44	0–6	1.3	19	2–6	3.1
10–0 to 10–11	34	0–4	0.8	9	2–4	3.1
11–0 to 11–11	36	0–6	1.1	11	2–6	3.5
12–0 to 12–11	23	0–6	1.1	7	2–6	3.4
13–0 to 13–11	9	0–2	0.9	4	2	2.0
14–0 to 14–11	18	0–4	0.6	4	2–4	2.5
15–0 to 15–11	2	2–12	7.0	2	2–12	7.0
Total.....	204	....	...	66		

Increases in mental age were found in 32.35 per cent of the cases. The average of the increases was 3.14 months of mental age. Sta-

tistically this may not appear to be a large change. However, in clinical practice we are dealing with individuals rather than with measures of central tendency. Since the range of increase was as great as twelve months (see Table II) an extension of the test beyond the first zero point seems warranted.

In the previous study it was pointed out that an extension of the test increases the number of items it is necessary to administer and requires the subject to respond to items of a higher level of difficulty, thereby increasing the possibility of fatigue and also affecting rapport. It should be remembered, however, that fatigue and loss of rapport do not detract in any way from the results obtained when the test is given under the standard procedure, since all increases in score are obtained after the first upper limit has been reached. Table III is given to indicate the range of testing when the suggested procedure is used.

TABLE III.—NUMBER OF ITEMS ADMINISTERED TO EACH CHILD FROM THE BASAL YEAR TO THE LAST ZERO POINT PRECEDED BY A SUCCESS

Age	Number of items											
	18	24	30	36	42	48	50	54	56	60	62	Average
7–0 to 7–11	0	2	3	1	2	0	0	0	0	0	0	32.25
8–0 to 8–11	2	7	6	8	5	2	0	0	0	0	0	32.60
9–0 to 9–11	2	6	4	9	14	6	0	3	0	0	0	37.77
10–0 to 10–11	2	2	6	7	13	3	0	0	0	0	1	37.29
11–0 to 11–11	0	5	4	11	6	4	0	2	2	1	1	39.50
12–0 to 12–11	0	1	3	4	4	7	0	3	1	0	0	42.61
13–0 to 13–11	0	0	2	2	2	2	0	1	0	0	0	40.66
14–0 to 14–11	0	0	1	4	4	7	1	1	0	0	0	43.77
15–0 to 15–11	0	0	0	1	0	0	0	0	0	0	1	49.00
Total (204).....	6	23	29	47	50	31	1	10	3	1	3	

On Form M in 13.7 per cent of the cases the subjects were required to respond to items at age levels from three to five years beyond their chronological ages. Even more significant is the fact that in 84.8 per cent of the cases, the subjects had to respond to items at age levels from three to seven years beyond their mental ages.

An analysis of the test items passed beyond the first year in which all items are failed shows that certain items are passed with considerable frequency. This is shown in Table IV.

TABLE IV.—TEST ITEMS PASSED AFTER FIRST ZERO POINT WAS REACHED AND FREQUENCY WITH WHICH THESE ITEMS WERE PASSED

Year level on Form M	Item number					
	1	2	3	4	5	6
VIII	0	0	0	0	0	1
IX	3	1	0	0	6	9
X	9	3	0	0	8	11
XI	0	2	0	0	0	5
XII	4	7	1	0	2	8
XIII	1	0	0	0	13	1
XIV	2	1	4	0	0	0
AA	0	0	0	0	0	1
SA-I	0	0	0	1	0	0

The rhymes, digits forward and reversed, word naming, sentence memory (year XI), response to picture (Messenger Boy) and problems of fact are among the items most frequently passed beyond the first zero point. This may indicate the possibility of inadequate scaling of the test for these items.

A comparison of the scatter on both Forms L and M indicates that a large percentage of retarded pupils increase their scores when the tests are extended. On Form L forty-two per cent of the scores were increased as compared with 32.35 per cent of the scores on Form M. The differences in per cent of increase was not found to be statistically significant (critical ratio 2.0). On both forms the average increase in months of mental age was found to be somewhat over three months. On both forms certain items were passed with considerable frequency after a year level of complete failures, indicating the possibility of inadequate scaling for these items. We may conclude that in the clinical examination of retarded children the test should be extended to at least a point where two levels of falures has been reached for it to be an adequate measure. This holds for both Forms L and M of the Revised Stanford-Binet Scale.

BOOK REVIEWS

W. H. KILPATRICK. *Selfhood and Civilization.* New York: The Macmillan Co., 1941, pp. 243.

This revision of the four lectures that Kilpatrick delivered during the Fall of 1938 on the V. Everit Macy Foundation of Teachers College, Columbia University, defends the thesis that the human personality is inherently a social product. No one is born a self. That status must be achieved and it can be achieved by man alone of all living organisms. The process of achieving selfhood is an extended one in time and involves various stages and degrees. Only in a social milieu can this self emerge, and it is indelibly marked by the culture that surrounds it. Selfhood, according to Kilpatrick, has been achieved: (*a*) When the individual is first able to distinguish himself clearly from others and does so by the use of appropriate pronouns, (*b*) when he recognizes himself as an agent, (*c*) when he has achieved a sense of time including a notion of the continuity of himself, and (*d*) when he has built a sense of conscious intent and accompanying accountability and responsibility.

When the self is so defined and it is then contended that the self is inherently a social product, the reviewer sees no possibility of argument. Distinguishing oneself from others would manifestly be impossible apart from a social environment. The concept of accountability or responsibility implies, as well, the presence of others.

Kilpatrick rightly places great stress upon the importance of language in the development of the "self" and of the personality. Without most of the idealistic presuppositions that the Hegelian makes, he arrives at the same conclusion; namely, that man is qualitatively different from the brute and that this qualitative difference is explicable chiefly in terms of man's use of language—a social phenomenon. The author thus feels that he has presented an alternative to both philosophical dualism and monism which "have for too long bedeviled psychology along with every other study of man." All of this is related in the last chapter to the current scene and the book concludes with a vigorous vilification of Hitler and the assertion that were he to win, America could have no security. The "great man" doctrine of historical interpretation is ubiquitous, at least by implication, throughout the argument. Only two pages are devoted to "our unsatisfactory economic system" (p. 186f). STEPHEN M. COREY.

University of Chicago.

S. Howard Bartley. *Vision, A Study of Its Basis.* New York: D. Van Nostrand Co., 1941, pp. 350.

During the past fifteen years there has been rapid advancement in the physiology of vision. Much of the experimentation was made possible by new electrical techniques. It is largely the data and theorizing accumulated during this recent period that furnish the material for this treatise. Several of these investigations were carried out by Dr. Bartley. Although the subtitle partially characterizes the contents, the major emphasis is on theoretical considerations arising from recent work on the physiology of vision. Consequently, as a treatise on vision it is highly specialized.

In dealing with visual experience two viewpoints are possible. "The one is known as mechanistic theory, elementalism or atomism; the other as configurationism, organismicism, Gestalt theory, or field theory." The author points out that physiology is almost totally atomistic. Consequently, most of the material presented here is analytical, although a few experiments involving field theory are given. It is granted that both approaches have their place although the elemental and the field concepts appear to be in sharp contrast. Can the two concepts be reconciled? "Only a radically different hypothesis than any yet used by nerve physiology can evolve the solution for which we seek." The author feels that the desirable thing is to be able to do in neural terms what has been done in perception itself; that is, to use our present physiological information logically in a field hypothesis. Possibly progress in this direction would be facilitated if the scientist could break away from the traditional view of "present physiological information."

The reader should expect neither a comprehensive nor an elementary handbook on the psychology of vision. At the outset the author limits the discussion to those visual phenomena about which contemporary physiology can make a direct contribution. "In the present treatment of vision all items which do not rather immediately contribute to a correlation between experience and neurophysiology will be laid aside." Within this scope the author has done an excellent job. Representative of the topics discussed are adaptation phenomena, repeated stimulation (flicker), and the electroretinogram. Basic relationships, such as those between brightness discrimination and visual acuity, are pointed out. Experimental data taken from the author's own work and from other sources also have been well

selected for developing a consistent exposition. The presentation is compact as well as orderly and will probably appeal most to the reader with extensive background. Psychologists will profit by what this physiologist has to offer.

MILES A. TINKER.
University of Minnesota.

KIMBALL YOUNG. *Personality and Problems of Adjustment.* New York: F. S. Crofts & Co., 1940, pp. 868.

With considerable skill, the author has combined the viewpoints and data of psychology, sociology and cultural anthropology in an endeavor to present an eclectic approach to the topic of personality and adjustment. The book is divided into three parts. A substantial portion of the more than three hundred pages contained in Part I is devoted to material that has but little relationship to the title. Sandwiched between chapters on constitutional foundation of behavior and theories of personality are long chapters on learning, language and thought. Part II, by far the best, contains excellent chapters on child adjustment, mental hygiene for college students, marital and occupational adjustment, juvenile delinquency, and mental disease. The contribution of religion, art, and avocation to personality integration, and the inter-relation of personality and culture are discussed in Part III. The appendix contains a useful outline for writing a case history.

As stated by the author in the prefatory note, this book was written for a rather unusual mixture of students However, it may be successfully used as a textbook for the usual college course on personality and mental hygiene by limiting readings to selected chapters.

JAMES D. PAGE.
Temple University.

ESTHER J. SWENSON. *Retroactive Inhibition; A Review of the Literature.* Minneapolis: University of Minnesota Press, 1941, pp. 59.

In recent years, experiments have yielded an increasing amount of evidence in support of the view that forgetting is the result of an "active process of interference," termed retroactive inhibition, rather than the result of a passive process of disuse. However, many of the applications of these experimental findings have been overlooked by the educational psychologist because of the impossibility of keeping up with the many publications in the field of general psychology along with those in the areas of education.

Therefore, the educator should welcome this monograph summarizing the experimental work on retroaction since S. H. Britt's publication in 1935 of a survey of the literature in the field. The implications of this research on retroaction are also discussed, special emphasis being given to the relationships of learning, retention, and transfer to educational problems. Thus, the monograph not only "suggests significant clues as to the 'why' of the vast educational waste due to forgetting," but also offers practical suggestions as to the "how" of remedial measures and further research.

An extensive bibliography of the recent literature is appended.

DAVID V. TIEDEMAN.

University of Rochester.

SERGE VORONOFF. *From Cretin to Genius.* New York: Alliance Book Corporation, 1941, pp. 281.

In this book Dr. Serge Voronoff, best known for his work in rejuvenation through the grafting of monkey glands, deals with what he chooses to call some "tantalizing enigmas" of the mind. Illustrative of such enigmas raised by the author are: Does the soul regulate our higher thoughts? Is it possible to attribute ordinary thoughts to cerebral matter? Is the brain the soul's instrument?

In actuality the book makes fairly interesting reading. There are chapters on creative processes in poets and writers, in composers, in scientists, and in mathematicians. There are also separate chapters on the struggles of geniuses similarly classified. In the struggle group the philosophers are added. There is a small chapter in the early part of the book called "From Cretin to Genius" and the last chapter is called "From Genius to Cretin."

Women may not be at all pleased to discover that, according to Voronoff, "woman is rarely possessed of high qualities equivalent to her husband's; her twenty-four chromosomes, inferior in quality, do not complement those of the genius." Members of Napoleon's family may be pleased to complete their geneology with some references of Voronoff. He informs us that Napoleon kept a bust of Julius Caesar in his study because he identified his genius with that of the great Roman, and then he informs us that "without knowing it, he may indeed have inherited, from Caesar, qualities ever-present in the germ cells of Caesar's male and female descendants, scattered throughout the world across the centuries." In like manner, perhaps even more

farfetched, does the author explain how it is possible for "a vulgar soul to give birth to a genius."

How he thinks about genius can be perceived from the following illustrative quotations: "Genius is an inborn autonomous faculty, independent of general mentality, and manifested by sudden inspiration"; or "the man of genius is a visionary; from birth he is provided by heredity with a marvelous aptitude, a sublime gift, independent of his general mentality, and related to a hypertrophied centre of the brain"; or "the wonderful qualities which genius displays are not necessarily associated with a superior intelligence; they express, rather, the functioning of the specialized centres genius inherited"; or "obviously there is a splitting of the personality in every genius." All of the foregoing illustrations indicate that Voronoff's treatment of genius is on the level that the study of animal intelligence was on before the days of Thorndike; namely, personal impressions and anecdotes.

The title of the book is perhaps somewhat misleading. It might lead some readers to think they will find in a volume so titled some references to genetic and hereditary studies of genius, or Terman's studies, or productivity studies in relationship to cultural and economic factors, or to motivational studies. But that is not what they will find in the book. They will find it pleasant and easy reading, with some amusing anecdotes and a few statements concerning Voronoff's own work, the nature of his own genius, and his own struggles.

The book ends with an elevating and ennobling thought to the effect that "we are not common clay, we are spirit incorporated in matter."

H. Meltzer.

Psychological Service Center, St. Louis, Missouri.